

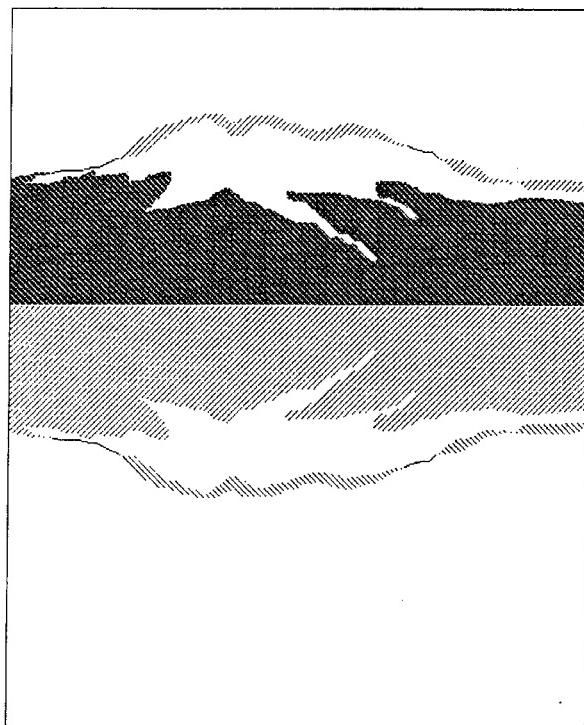
ISSN 0031-8949
(ISSN 0281-1847)
ISBN 91-87308-21-5

Proceedings of

16th Nordic Semiconductor Meeting

Laugarvatn, Iceland, June 12–15, 1994

Editors: H. Gíslason
V. Guðmundsson



Recognized by the European Physical Society



Physica Scripta

19990204 024

Vol. T54 1994

DISSEMINATION STATEMENT A
Approved for public release; Distribution Unlimited

A6 F99-05-0848

Physica Scripta

Published jointly by

The Royal Danish Academy of Sciences and Letters in cooperation with the Danish Physical Society

The Delegation of the Finnish Academies of Science and Letters in cooperation with the Finnish Physical Societies

The Icelandic Scientific Society in cooperation with the Icelandic Physical Society

The Norwegian Academy of Science and Letters in cooperation with the Norwegian Physical Society

The Royal Swedish Academy of Sciences in cooperation with the Swedish Physical Society

Editors

General and cross-disciplinary physics

***P. L. Christiansen** (theor)

M. Høgh Jensen (theor)

K. Mørk (theor)

P. Pierański (exp)

High energy physics

K. Enqvist (theor astropart phys)

E. Lillestøl (exp)

U. Lindström (theor)

Nuclear physics

G. B. Hagemann (exp)

***D.-O. Riska** (theor)

Atomic, molecular and optical physics

Dž. Belkić (theor & exp)

L. J. Curtis (theor & exp)

J. Javanainen (theor & exp)

***I. Martinson**

A. Rosén (theor & exp)

V. P. Shevelko (theor)

S. Svanberg (exp)

G. Werth

Plasma physics

***H. L. Pécseli** (theor & exp)

***L. Stenflo** (theor)

M. Y. Yu (theor)

Condensed matter physics and material sciences

P. Apell (theor)

***H. P. Gíslason** (exp)

H. G. Grimmeiss (exp)

***V. Guðmundsson** (theor)

P. C. Hemmer (theor)

T. Jøssang (theor & exp)

P. E. Lindelof (exp)

***R. Nieminen** (theor & exp)

***T. Riste** (exp)

W. Salaneck (exp)

T. I. Sigfusson (exp)

Geophysics, astronomy and astrophysics

S. J. Johnsen

J. M. Knudsen (exp)

*Member of the editorial board

Editorial Office

Scientific Editors

Anders Bárány (executive)

Roger Wäppling

Agneta Seidel

Production Editors

Åsa von Krusenstjerna

Katarina Lundin

Manuscripts

Shall be sent in triplicate – one original and two copies – to any of the editors. For full addresses see cover page 3.

Offprints

25 offprints of each paper will be supplied free of charge. Additional copies may be ordered at cost price.

Subscriptions

The Royal Swedish Academy of Sciences

Physica Scripta

Box 50005

S-104 05 Stockholm

Sweden

Subscription price

For 1994 (vols 49, 50 and T-volumes) the subscription price is DEM 1575. For subscribers in the Nordic Countries SEK 4725. For subscribers in USA, Canada and Mexico US \$900.

Back numbers and single issues

Can be ordered separately. Price information on request.

Physica Scripta is delivered to some institutions in the former Soviet Union through a program initiated and administered by the European Physical Society. This program is funded by INTAS, the International Association for the promotion of cooperation with scientists from the independent States of the former Soviet Union. Members of INTAS are the European Union, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, the Netherlands, Portugal, Spain, United Kingdom, Sweden and Switzerland. (INTAS, Rue du Luxembourg 14A, 1040 Bruxelles, Belgique).

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

			3. REPORT TYPE AND DATES COVERED Conference Proceedings
1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE 10 January 1995	4. TITLE AND SUBTITLE 16th Nordic Semiconductor Meeting	
6. AUTHOR(S) Conference Committee		5. FUNDING NUMBERS F6170894W0558	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Iceland Dunhagi 3 Reykjavik IS-107 Iceland			8. PERFORMING ORGANIZATION REPORT NUMBER N/A
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) EOARD PSC 802 BOX 14 FPO 09499-0200			10. SPONSORING/MONITORING AGENCY REPORT NUMBER CSP 94-1024
11. SUPPLEMENTARY NOTES 2 volumes - 1 abstracts, 1 proceedings			
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.			12b. DISTRIBUTION CODE A
13. ABSTRACT (Maximum 200 words) The Final Proceedings for 16th Nordic Semiconductor Meeting, 12 June 1994 - 15 June 1994			
14. SUBJECT TERMS EOARD			15. NUMBER OF PAGES 493
			16. PRICE CODE N/A
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

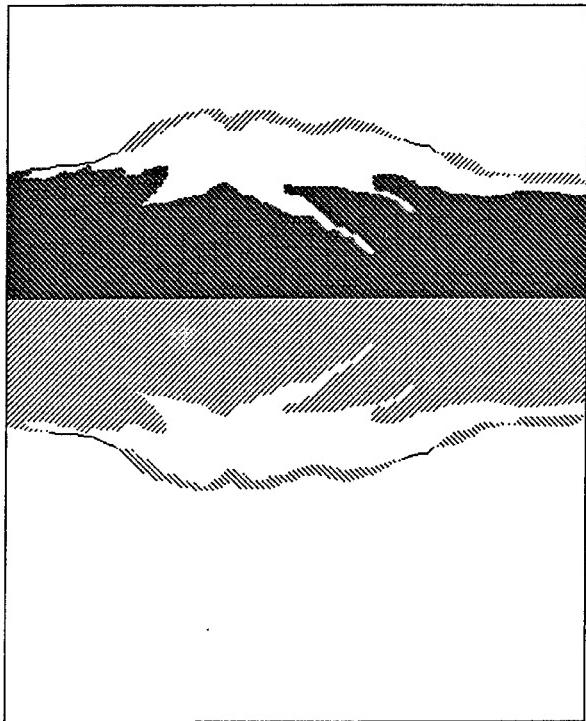
ISSN 0031-8949
(ISSN 0281-1847)
ISBN 91-87308-21-5

Proceedings of

16th Nordic Semiconductor Meeting

Laugarvatn, Iceland, June 12–15, 1994

Editors: H. Gíslason
V. Guðmundsson



Recognized by the European Physical Society



Physica Scripta

Vol. **T54** 1994

Physica Scripta

Published jointly by

The Royal Danish Academy of Sciences and Letters in cooperation with the Danish Physical Society

The Delegation of the Finnish Academies of Science and Letters in cooperation with the Finnish Physical Societies

The Icelandic Scientific Society in cooperation with the Icelandic Physical Society

The Norwegian Academy of Science and Letters in cooperation with the Norwegian Physical Society

The Royal Swedish Academy of Sciences in cooperation with the Swedish Physical Society

Editors

General and cross-disciplinary physics

***P. L. Christiansen** (theor)

M. Høgh Jensen (theor)

K. Mørk (theor)

P. Pierański (exp)

High energy physics

K. Enqvist (theor astropart phys)

E. Lillestöl (exp)

U. Lindström (theor)

Nuclear physics

G. B. Hagemann (exp)

***D.-O. Riska** (theor)

Atomic, molecular and optical physics

Dž. Belkić (theor & exp)

L. J. Curtis (theor & exp)

J. Javanainen (theor & exp)

***I. Martinson**

A. Rosén (theor & exp)

V. P. Shevelko (theor)

S. Svanberg (exp)

G. Werth

Plasma physics

***H. L. Pécseli** (theor & exp)

***L. Stenflo** (theor)

M. Y. Yu (theor)

Condensed matter physics and material sciences

P. Apell (theor)

***H. P. Gislason** (exp)

H. G. Grimmeiss (exp)

***V. Guðmundsson** (theor)

P. C. Hemmer (theor)

T. Jøssang (theor & exp)

P. E. Lindelof (exp)

***R. Nieminen** (theor & exp)

***T. Riste** (exp)

W. Salaneck (exp)

T. I. Sigfusson (exp)

Geophysics, astronomy and astrophysics

S. J. Johnsen

J. M. Knudsen (exp)

*Member of the editorial board

Editorial Office

Scientific Editors

Anders Bárány (executive)

Roger Wäplling

Agneta Seidel

Production Editors

Åsa von Krusenstjerna

Katarina Lundin

Manuscripts

Shall be sent in triplicate – one original and two copies – to any of the editors. For full addresses see cover page 3.

Offprints

25 offprints of each paper will be supplied free of charge. Additional copies may be ordered at cost price.

Subscriptions

The Royal Swedish Academy of Sciences

Physica Scripta

Box 50005

S-104 05 Stockholm

Sweden

Subscription price

For 1994 (vols 49, 50 and T-volumes) the subscription price is DEM 1575. For subscribers in the Nordic Countries SEK 4725. For subscribers in USA, Canada and Mexico US \$900.

Back numbers and single issues

Can be ordered separately. Price information on request.

Physica Scripta is delivered to some institutions in the former Soviet Union through a program initiated and administered by the European Physical Society. This program is funded by INTAS, the International Association for the promotion of cooperation with scientists from the independent States of the former Soviet Union. Members of INTAS are the European Union, Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, the Netherlands, Portugal, Spain, United Kingdom, Sweden and Switzerland. (INTAS, Rue du Luxembourg 14A, 1040 Bruxelles, Belgique).

Contents

Preface	5
Defects	
Configurational instabilities at isoelectronic centres in silicon. <i>G. Davies</i>	7
Lithium-gold-related defect complexes in n-type silicon. <i>E. Ö. Sveinbjörnsson, S. Kristjánsson and H. P. Gislason</i>	12
Conversion electron Mössbauer spectroscopy study of iron silicide films grown by MBE. <i>M. Fanciulli, G. Weyer, H. von Känel and N. Onda</i>	16
Deep centre photoluminescence in nitrogen doped ZnSe. <i>I. S. Hauksson, S. Y. Wang, J. Simpson, K. A. Prior, B. C. Cavenett, W. Liu and B. J. Skromme</i>	20
Excitation of the 4f-electron of Pr ³⁺ in GaAs : Pr and Al _x Ga _{1-x} As : Pr. <i>P. L. Thée, Y. K. Yeo, R. L. Hengehold and G. S. Pomrenke</i>	24
Passivation of shallow and deep levels by lithium in GaAs. <i>T. Egilsson, B. Yang and H. P. Gislason</i>	28
Effect of the interatomic Si-Si-potential on vacancy production during ion implantation of Si. <i>K. Nordlund, J. Keinonen and A. Kuronen</i>	34
Design and fabrication	
Surface morphology study of titanium silicide formed on polycrystalline silicon. <i>T. E. Karlin, W. Kaplan and S. Zhang</i>	38
A fine pattern GTO thyristor fabricated using a self-aligned process. <i>M. Bakowski, U. Gustafsson, H. Elderstig, W. Kaplan, P. Norlin and U. Wennström</i>	42
Application of III-V-semiconductor based heterojunction bipolar transistors towards multi-Gbit/s 4 : 1 multiplexer. <i>G. Schuppener, B. Willén, M. Mokhtari and H. Tenhunen</i>	46
Multichamber processor for small semiconductor laboratories – the first results. <i>R. Punkkinen, H. Ihantola, K. Jokinen, T. Kuusela, H. Arvela, H.-P. Hedman and V. Saikku</i>	51
Design optimization of negative charge pumps for IGBT driver ICs. <i>M. H. Åberg</i>	54
Electrical devices	
Processing and characterisation of an etched groove permeable base transistor on 6H-SiC. <i>C. Fröjd, G. Thungström, S. Hatzikonstantinou, H.-E. Nilsson and C. S. Petersson</i>	56
The influence of emitter properties on the heat generation in SiC and Si PIN diodes under forward conduction. <i>O. Tornblad, B. Breitholtz, M. Östling and U. Lindefelt</i>	60
High-voltage silicon carbide rectifiers – results of experiments and simulation. <i>L. P. Ramberg, S. Savage, U. Gustafsson and A. Schöner</i>	65
Metallizations of InP based on transition metals. <i>T. Clausen and O. Leistikko</i>	68
Impedance field and microwave power generation in InP diodes. <i>V. Gružinskis, E. Starikov and P. Shiktorov</i>	71
Contacts to monocrystalline n- and p-type silicon by wafer bonding using cobalt disilicide. <i>G. Thungström, C. Fröjd, P. Svedberg and C. S. Petersson</i>	77
Low dimensional systems	
Magneto luminescence of As-grown InAs/InP quantum well islands. <i>H. Sigg, P. Christianen, R. Houtré and A. Rudra</i>	81
Initial strain relaxation and optical quality in lattice mismatched InGaAs/GaAs single quantum wells. <i>S. M. Wang and T. G. Andersson</i>	84
Electron correlation effects in quantum dots. <i>A. Matulis</i>	88
Enhancement of the g-factor and spin-density wave state in a confined 2DEG in the quantum Hall regime. <i>V. Guðmundsson and G. Pálsson</i>	92
Dissipation in the quantum Hall effect by transverse circulation of electrons. <i>E. B. Hansen</i>	96
Spin resonances determination of the effective g-factor of electrons in low dimensional (GaIn)As/InP structures. <i>B. Kowalski, P. Omling, B. K. Meyer, D. M. Hofmann, V. Härlle, F. Scholz and P. Sobkowicz</i>	100
Mesoscopic systems	
Characteristic potentials for mesoscopic rings threaded by an Aharonov-Bohm flux. <i>M. Büttiker</i>	104
Raman scattering from a circular quantum dot. <i>A. G. Mal'shukov, A. Brataas and K. A. Chao</i>	111
Local field calculation for a spherical semiconductor quantum dot with parabolic confinement. <i>O. Keller and T. Garm</i>	115
Resonant tunneling: from model Hamiltonian to modern electronic devices. <i>K. A. Chao, M. Willander and Yu. M. Galperin</i>	119
Aharonov-Bohm effect in a quantum ring with strong electron-electron correlations. <i>I. V. Krive, R. I. Shekhter, S. M. Girvin and M. Jonson</i>	123
Modelling of devices	
Unified capacitance modelling of MOSFETs. <i>O. G. Johannessen, T. A. Fjeldly and T. Ytterdal</i>	128
GaAs/AlGaAs quantum well infrared photodetector arrays for thermal imaging applications. <i>L. Lundqvist, J. Y. Andersson and J. Borglind</i>	131
A compact model for the cutoff frequency in high speed bipolar transistors. <i>M. Andersson, P. Kuivalainen, Z. Xia, H. Pohjonen and H. Ronkainen</i>	136
Extrinsic versus intrinsic models for FETs. <i>T. Ytterdal, T. A. Fjeldly and K. Lee</i>	139
The effect of using different transport models in computer simulations of the permeable-base transistor. <i>H.-E. Nilsson, U. Sannemo, A. Koel, F. Masszi and C. S. Petersson</i>	141
Monte Carlo simulation of hot carrier noise in short n ⁺ nn ⁺ diodes. <i>V. Gružinskis, E. Starikov and P. Shiktorov</i>	146
Effect of partial ionization and the characteristics of lateral power diamond MESFETs. <i>K. J. Grahn, P. Kuivalainen and S. Eränen</i>	151
Quantum corrections to the threshold voltage of short channel MOSFETs. <i>P. Kuivalainen</i>	154
Physical modelling of vertical DMOS power transistors for circuit simulation. <i>M. Andersson and P. Kuivalainen</i>	157
A semi-analytic model of the permeable base transistor. <i>H.-E. Nilsson, U. Sannemo and C. S. Petersson</i>	159
Optical devices	
Interferometric, low thermal mass IR-absorber for thermal infrared detectors. <i>P. Eriksson, J. Y. Andersson and G. Stemme</i>	165
UV-sensitive photodetectors based on metal-semiconductor contacts on 6H-SiC. <i>C. Fröjd, G. Thungström, H.-E. Nilsson and C. S. Petersson</i>	169
Large area GaInAsP and GaInP solar cells for space applications. <i>K. Smekalin, K. Tappura and J. Lammasniemi</i>	172
Fitting of the solar cell I/V-curve to the two diode model. <i>A. Hovinen</i>	175
Development of materials for blue/green light emitters. <i>K. Rakennus, P. Uusimaa, K. Smekalin, P. Savolainen and M. Pessa</i>	177

Photonics

Ultrafast nonlinear optics in GaAs/AlGaAs quantum wells. <i>J. M. Hvam, D. Birkedal, V. G. Lyssenko, J. Erland and C. B. Sørensen</i>	181
Modulation- and transmission-ellipsometric characterization of semiconductor heterostructures. <i>K. B. Ozanyan, T. Worren and O. Hunderi</i>	187
All-optical bistability in luminescence of thin CdS films. <i>B. Ullrich and T. Kobayashi</i>	191

Process technology

Selective-area MOVPE for InP-based optoelectronic components. <i>P. Tidemand-Petersson, O. Albrektsen and J. Salzman</i>	194
Interdiffusion and phase formation during thermal processing of Co/Ti/Si(100) structures. <i>J. Cardenas, S. Hatzikonstantinidou, S.-L. Zhang, B. G. Svensson and C. S. Petersson</i>	198
Laser ablation deposition as a preparation method for electronic materials. <i>S. Leppävuori</i>	202
Growth of silicon carbide on (100) silicon substrates by molecular beam epitaxy. <i>V. M. Airaksinen, J. Kaitila, H. Niemi, J. Lahtinen and J. Saarilahti</i>	205
Growth and characterization of compositionally graded, relaxed $\text{Si}_{1-x}\text{Ge}_x$. <i>A. N. Larsen, J. L. Hansen, R. S. Jensen, S. Y. Shiryaev, P. R. Østergaard, J. Hartung, G. Davies, F. Jensen and J. W. Petersen</i>	208
RBS channeling spectroscopy of Ge implanted epitaxial $\text{Si}_{1-x}\text{Ge}_x$ layers. <i>J. Saarilahti, Z. Xia, H. Ronkainen, P. Kuivalainen and I. Suni</i>	212
As capping of MBE-grown compound semiconductors; novel opportunities to interface science and device fabrication. <i>J. K. Grepstad, H. Husby, A. Borg, B.-O. Fimland, R. W. Bernstein and R. Nyholm</i>	216
Process optimisation and characterisation of PBT structures. <i>S. Hatzikonstantinidou, H.-E. Nilsson, C. Fröjdh and C. S. Petersson</i>	226
A flip chip process based on electroplated solder bumps. <i>J. Salonen and J. Salmi</i>	230
Low temperature silicon epitaxy in a single-wafer RTP reactor with microwave heating. <i>S.-L. Zhang and R. Buchta</i>	234
The Au/Si(111) system studied by optical second-harmonic generation. <i>K. Pedersen and P. Morgan</i>	238
Fabrication of nanostructures using MBE and MOVPE. <i>J. Ahopelto, H. K. Lipsanen, M. Sopanen, T. Koljonen, T. Tuomi, V. M. Airaksinen, J. Sinkkonen and E. Sirén</i>	241
Pulsed laser ablation deposition of CuInSe_2 and $\text{Culn}_{1-x}\text{Ga}_x\text{Se}_2$ thin films. <i>J. Levoska, S. Leppävuori, F. Wang, O. Kusmartseva, A. E. Hill, E. Ahmed, R. D. Tomlinson and R. D. Pilkington</i>	244
A study of dual conductance response to carbon monoxide of CdS and $\alpha\text{-SnWO}_4$ thin films. <i>J. L. Solis, V. Golovanov, V. Lantto and S. Leppävuori</i>	248

Recent advances in theory

A cluster approach for modelling of surface characteristics of stannic oxide. <i>T. S. Rantala, V. Lantto and T. Rantala</i>	252
The ionisation process of α particles in mesoscopic structures: simulation by Monte Carlo method. <i>K. Tarnay, F. Masszi, T. Kocsis, A. Poppe and L. Kiss</i>	256
A theoretically accurate mobility model for semiconductor device drift-diffusion simulation. <i>E. Velmre, A. Udal, T. Kocsis and F. Masszi</i>	263
Different methods of noise reduction in Monte Carlo simulations of a Schottky diode. <i>H.-E. Nilsson, U. Sannemo and C. S. Petersson</i>	268

Solid state materials

Cobalt disilicide (CoSi_2) Schottky contacts to 6H-SiC. <i>N. Lundberg and M. Östling</i>	273
The Pt/Si(111) interface and the properties of thin Pt layers on Si. <i>P. Morgen, B. Jørgensen and J. Gordon</i>	278
SiC – a semiconductor for high-power, high-temperature and high-frequency devices. <i>E. Janzén, O. Kordina, A. Henry, W. M. Chen, N. T. Son, B. Monemar, E. Sörman, P. Bergman, C. I. Harris, R. Yakimova, M. Tuominen, A. O. Konstantinov, C. Hallin and C. Hemmingsson</i>	283
Thermal oxidation of n- and p-type 6H-silicon carbide. <i>C.-M. Zetterling and Östling</i>	291
Strain relaxation in epitaxial $\text{Si}_{1-x}\text{Ge}_x$ layers during some silicidation processes. <i>O. Nur, M. R. Sardela Jr., H. H. Radamson, M. Willander, G. V. Hansson and S. Hatzikonstantinidou</i>	294
An investigation of the stability of copper germanide thin films in the presence of Si and SiO_2 . <i>J. P. Doyle, B. G. Svensson, M. O. Aboelfotoh and J. Hudner</i>	297
Properties of iron silicide contacts to n- and p-type silicon. <i>U. Erlesand and M. Östling</i>	300
List of participants	305

Preface

Some 30 years ago an informal meeting of the few Nordic specialists in semiconductor physics marked the beginning of what has become a biannual meeting of some hundred physicists and physics students from all the Nordic countries. The 16th Nordic Semiconductor Meeting took place at Laugarvatn, Iceland, June 12–15, 1994.

As a regional meeting the Nordic Semiconductor meeting has three characteristic features all of which distinguish it from more traditional international meetings in the field. First, it has the purpose of promoting Nordic cooperation in the international field of semiconductor physics. Research in the fields of advanced science and technology in the Nordic countries is likely to benefit from joining national forces before participating in the increasing European integration. Second, there is an unusually large fraction of graduate students amongst the participants of the Nordic Semiconductor Meeting. In fact, attending this conference is traditionally a part of the graduate program in semiconductor physics and technology. The Nordic Semiconductor Meeting is often the first conference of international character that graduate students attend in order to present a paper or poster. Third, there is an interdisciplinary quality of the meeting which is normally not the case for meetings of this size. In particular, the number of professional scientists from industry is comparable to the number of their academic colleagues. This is important for both groups, but perhaps the graduate students benefit most from presenting their results to both groups.

The 16th Nordic Semiconductor Meeting, the first one in this series held in Iceland, attracted 129 active participants. The scientific programme was divided in twelve oral sessions. A novelty of this meeting was the emphasis on more fundamental physics in one of the two parallel sessions but more applied topics in the other, although the distinction was sometimes a matter of predilection. A poster session including both basic and applied physics was also organized. Most of the oral sessions included an invited lecture. The invited speakers were all of high international class, five of them working in the Nordic countries, Sami Fransila, Finland, Jostein Grepstad, Norway, Jørn Hvam, Denmark, Erik Janzen and Lars Samuelson, Sweden. The other five represented a wider geographical spread, Klaus von Klitzing and Detlef Heitmann, Germany, Gordon Davies, United Kingdom, Markus Büttiker and Chris Palmstrøm, U.S.A. Attendees from China, Japan, Switzerland, the Netherlands and Lithuania also participated in the conference. In addition to the invited lectures some 100 oral papers and 25 posters were contributed.

Another novelty of the conference is the fact that the proceedings of the conference are being published in a refereed journal. These proceedings contain all the invited and contributed papers the authors of which complied with the deadline of submission of the manuscripts. The editors paid special attention to prompt publication of the proceedings in order to promote the actuality of the results presented at the conference. Therefore, the deadline was strict, all of the papers were refereed during the conference. Changes suggested by the referees were either made at Laugarvatn or within three weeks from the conference. We are grateful to the international crowd of session chairmen who assumed the task of refereeing the papers, either themselves or with the help of colleagues. Without their impressive qualifications this procedure would not have been as reliable as the quality of the papers deserved. We also want to thank the editorial staff of *Physica Scripta* for their help and cooperation.

It is our hope that the 16th Nordic Semiconductor Meeting succeeded in keeping the tradition of a popular conference series at the same time as modifying slightly the emphasis which may strengthen future meetings. Time will tell. In the meantime, we thank all the participants for their contributions. We are grateful to the sponsors listed below. Their support made the conference possible. Last, but not least, we acknowledge the work of Gerlinde Xander and all the students and co-workers who attended to countless details.

Reykjavík 18.7. 1994

Hafliði Pétur Gíslason
Viðar Guðmundsson

Acknowledgments

The organizers of the conference would like to express their sincere gratitude to the following sponsors of the **16th Nordic Semiconductor Meeting**:

NorFA

NORDITA

Raunvísindastofnun Háskóla Íslands (Science Institute, University of Iceland)

Háskóli Íslands (University of Iceland)

Eimskip hf.

Örtölvutækni

Smith & Norland

Hitaveita Reykjavíkur (Reykjavík District Heating)

Póstur og sími (Post and Telecommunication Services in Iceland)

Rafmagnsveitir Ríkisins (The State Electric Power Works)

Support to young Scandinavian scientists attending the conference:

Clara Lachmanns Fond

Support to students attending the conference:

NorFA

Iðnaðarráðuneytið (Ministry for Industry)

Marel

Landsvirkjun (National Power Company)

Support to Baltic participants:

Nordic Council of Ministers

Menntamálaráðuneytið (Ministry for Culture)

We additionally wish to thank the United States Air Force European Office of Aerospace Research and Development for its contribution to the success of this conference.

Configurational Instabilities at Isoelectronic Centres in Silicon

Gordon Davies

Physics Department, King's College London, Strand, London WC2R 2LS, UK

Received May 9, 1994; accepted June 15, 1994

Abstract

When an exciton is bound on an isoelectronic centre, one particle (electron or hole) may be severely localised on the centre. In this paper attention is drawn to the importance of the lattice relaxations stimulated by this localised charge density. Two examples are discussed in detail for centres where the exciton binding is produced primarily by the relaxation. It is shown that Jahn-Teller theory applied to the tightly bound particle allows a very precise understanding of the effects of external perturbations on the bound exciton.

1. Introduction

Luminescence from isoelectronic centres has been investigated for over three decades [1]. One motivation has been the very high quantum efficiency of the centres, which results from the centres having the same valence properties as the host atoms – recombination of an exciton leaves the centre in a “particle-free” state, in contrast to the recombination of an exciton at a charged centre where the additional charge may be excited in a non-radiative (e.g. Auger) process. However, despite the considerable studies of these centres, it is still not clear how their excited states are bound on the centres [2]. The prevalent view is that one particle (electron or hole) is bound by a short-range potential and the second (hole or electron) is trapped in the Coulomb field of the first [3]. The second particle may then have a quasi-Rydberg series of excited states, and such a series has been observed at some centres [4]. It has long been recognised that modifications to the exciton structure occur when the symmetry of the centre is lower than tetrahedral, and the axial properties may be described by a local strain field, as discussed by Gislason *et al.* [5]. In this paper we introduce a further concept where the local strain is produced by the presence of the exciton. We will discuss two examples of isoelectronic centres in silicon which illustrate some of the consequences of the lattice relaxations.

2. A simple description

The Hopfield–Thomas–Lynch model of bound excitons [3] assumes that one particle (say the electron) is spatially located on the core of the isoelectronic centre, the binding being derived possibly from the change in electron affinity at the centre. The second particle (the hole) is then bound by the Coulomb field of the electron. Since the electron is highly localised, its charge density is approximately point-like on the scale of the hole orbital, and the hole orbits in an approximately effective-mass-like state.

At this stage we introduce an ingredient which does not seem to have been explicitly discussed before in the context

of binding the exciton. The high charge density of the localised electron will force a local lattice relaxation. For a simple general discussion we consider first only hydrostatic deformations. A uniform hydrostatic deformation $(V - V_0)/V_0$ of the correct sign reduces the energy of the electron by $A(c_{11} + 2c_{12})(V - V_0)/3V_0$ where A is the change in electron energy per unit hydrostatic pressure. Consequently the energy of the electron is reduced if the lattice deforms over the volume V_0 occupied by the electron orbital. In the lattice-continuum approximation, the elastic energy required for this deformation is $(c_{11} + 2c_{12})(V - V_0)^2/6V_0$, assuming it is a uniform deformation. The total energy E of the electron plus the deformed lattice is

$$E = (c_{11} + 2c_{12})[(V - V_0)^2/6V_0 + A(V - V_0)/3V_0], \quad (1)$$

and is minimised (for a given V_0) at $V - V_0 = -A$, when

$$E_{\min} = -(c_{11} + 2c_{12})A^2/6V_0. \quad (2)$$

The energy saved by the deformation increases as V_0 decreases, so that the lattice relaxation tends to increase the localisation of the electron. However, the limit $V_0 \rightarrow 0$ is prohibited by the increasing kinetic energy of the localised electron.

A useful qualitative picture is that the energy reduction is produced by the local strain acting on the one electron, but to generate that strain involves deforming n chemical bonds within the extent of the electron orbital; the smaller n and the larger the strain that can be produced. Consequently self-binding effects can only occur when at least one particle is in a small orbital.

To illustrate the potential importance of lattice relaxations for a real centre, suppose that the isoelectronic centre has one substitutional impurity atom, and that the electron is sufficiently localised that the lattice deformation is limited to the four nearest neighbour atoms. We label these atoms as a, b, c, d at coordinates $\bar{1}\bar{1}1, \bar{1}\bar{1}\bar{1}, \bar{1}1\bar{1}, 1\bar{1}\bar{1}$. The hydrostatic breathing mode with a coordinate Q defined by

$$Q = (-x_a + x_b + x_c - x_d + y_a - y_b + y_c - y_d + z_a + z_b - z_c - z_d)/\sqrt{12} \quad (3)$$

has an effective mass equal to the mass M of one atom. A hydrostatic stress s produces a movement $Q = 2ls/(c_{11} + 2c_{12})$ where l is the interatomic spacing. This movement increases the elastic energy by $\frac{1}{2}M\omega^2 Q^2$, but decreases the electron energy by $3A(c_{11} + 2c_{12})Q/2l$. Stability is reached when the total energy is reduced by

$$\Delta E = 9A^2(c_{11} + 2c_{12})^2/8Ml^2\omega^2, \quad (4)$$

where ω is the angular frequency of the breathing mode. We take $A = 3 \text{ meV/GPa}$ and $\omega = 3.2 \times 10^{13} \text{ s}^{-1}$ (both values derived below from experiment), and use the perfect lattice values for the interatomic spacing l (0.234 nm) and the

elastic constants $c_{11} + 2c_{12}$ (298 GPa). The calculated value of $\Delta E = 55$ meV shows that the lattice relaxation produced by the electron can be a considerable contribution to its binding energy, even for the relatively small value of the deformation potential A used here, as long as the electron is sufficiently localised.

All this discussion can be rephrased for the alternative case of a hole-attractive centre, for which the electron is the effective-mass-like particle, and we will consider examples of both types of centre. We note that when the electron and hole recombine, the high charge density of the localised particle is removed, so that the source of the lattice relaxation disappears. The properties of the vibronic sideband produced by the luminescence transition give us information on the lattice relaxation, as we see next.

3. The "ABC" centre – an electron trap

One well-known centre in silicon, referred to as the ABC centre, produces luminescence with zero-phonon lines at 1122.3 meV at low temperature, corresponding to an exciton binding energy (measured from the free exciton level) of 32.9 meV [6]. Its chemical origin is not certain, except that the centre contains nitrogen [7]. The luminescence band, Fig. 1, has a vibronic sideband with a centroid 35 ± 3 meV below the zero-phonon line – from standard vibronic theory [8] this energy is the lattice relaxation ΔE of eq. (4). The vibronic sideband has a dominant peak at $\omega = 3.2 \times 10^{13} \text{ s}^{-1}$ (the value used in Section 2) which is not a critical point in the phonon distribution [9], and phonons with a wide range of wavevectors are present in the phonon sideband (Fig. 1). This spread in wavevectors implies that we are dealing with a transition involving a highly localised state.

Since we are interested in the effects of lattice relaxations on the excited states, we consider the effects of externally applied, controlled uniaxial stresses on the zero-phonon optical transitions. Sample data are shown in Fig. 2 for the effects of $\langle 001 \rangle$ compressions. The points [10] show the effects of stress as measured at 20 K (to give adequate population of the higher excited states), and the lines are a calculated fit (taking into account the effects also of $\langle 111 \rangle$ and $\langle 110 \rangle$ compressions). The fit is based on a model in which an electron is tightly bound in a non-degenerate orbital state, and the hole is weakly bound. The 6 conduction band minima, denoted $X, \bar{X} \dots \bar{Z}$, form states transforming as A_1, E , and T_2 in the T_d point group of a substitutional atom. For an electron-attractive centre we expect the A_1 state to be lowest in energy (as for a substitutional donor), and the fit in Fig. 2 (made with this assumption) shows this to be the case for the ABC centre. Zeeman measurements [6] on the luminescence emitted by the bound exciton show that it is trapped on the centre in a trigonal symmetry. In trigonal symmetry the A_1 electron state may be perturbed in first order by the hydrostatic component ($s_{xx} + s_{yy} + s_{zz}$) of the stress. Group theory also allows perturbation by a symmetry-maintaining compression along the trigonal axis, ($s_{xy} + s_{yz} + s_{zx}$). However, sheer stresses have no effect on the conduction band minima of silicon, and so the electron cannot be perturbed by any symmetry lowering stresses. The result is that no information can be derived about the trigonal distortion from the effects of stress on the tightly-

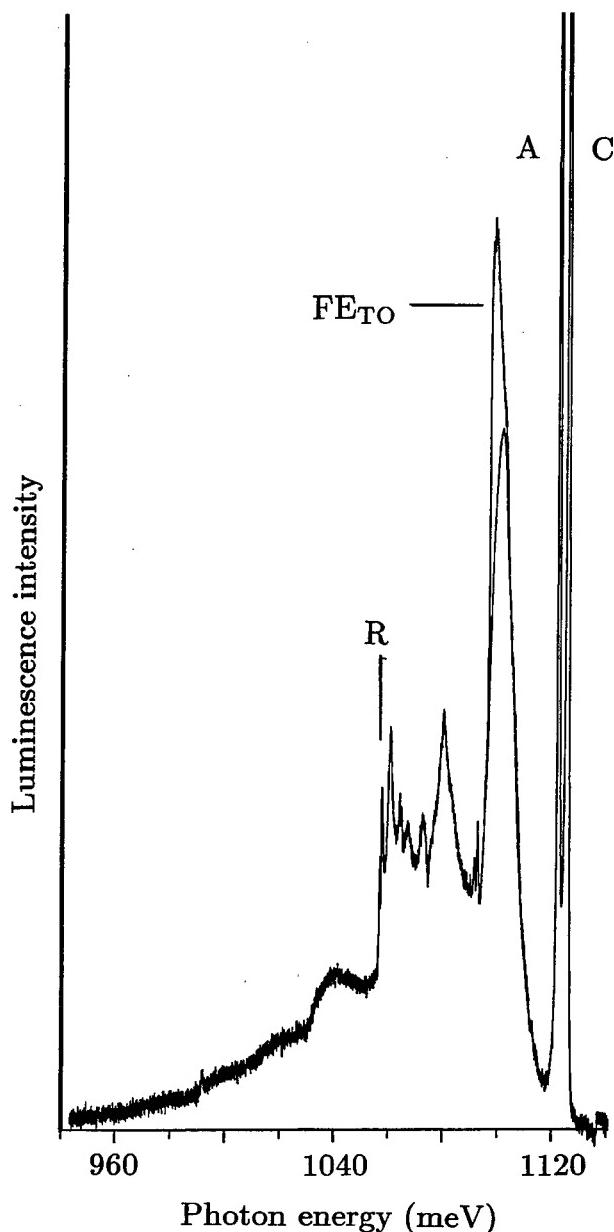


Fig. 1. Photoluminescence recorded at 20 K from the "ABC" isoelectronic centre in silicon. Zero-phonon lines A and C produce the uniaxial stress splittings shown in Fig. 2. Superimposed on the acoustic phonon sideband is emission from the free excitons (with the involvement of one transverse optic phonon), and the baseline for this emission is indicated. (This baseline is confirmed by the bandshape of the ABC system at low temperature). The extent of the one-phonon sideband is shown by the sharp cutoff at the Raman phonon, indicated by "R". Experimental data by M. Zafar Iqbal [10].

bound electron, even though it may be orbiting in a highly deformed region of the crystal.

The trigonal symmetry splits the $m_j = \frac{3}{2}$ valence band states [5], and is primarily responsible for the 3 meV splitting observed at zero stress. All the stress-induced splitting on Fig. 2 derives from the perturbation of the hole by the stresses. The fit is achieved with the deformation potentials of the hole being within 25% of those of the valence band states [11]; the small change would be consistent with the hole being relatively de-localised. Small deviations could occur simply from local changes in the elastic constants as well as from the modifications of the hole's properties produced by its being bound. The fit on Fig. 2 uses a hydrostatic response for all the states of 3 meV/GPa, and the fact

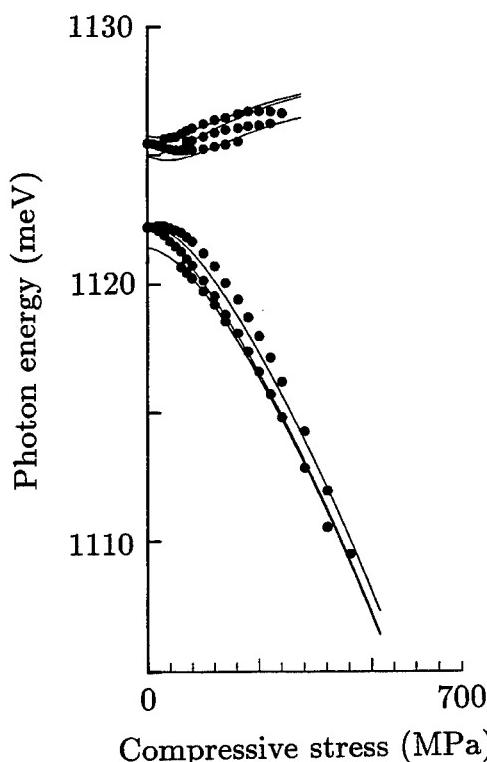


Fig. 2. Effects of uniaxial stresses on the A (1122 meV) and C (1125 meV) zero-phonon lines, as recorded at 20K for compressions along the $\langle 001 \rangle$ axis. The lines are the calculated effects of stresses on those transitions which are predicted to be optically allowed. The B line, with zero-stress value near 1121 meV, is induced by the stresses. Experimental data by M. Zafar Iqbal [10].

that the same value can be used for all the transitions is consistent with the common parentage of all the states.

The 3 meV splitting at zero applied stress is equivalent to a compression along $\langle 111 \rangle$ of about 94 MPa, and presumably reflects the $\langle 111 \rangle$ axis of the core of the centre, as seen by the relative delocalised hole. Paradoxically, in this example the tightly bound particle (the electron) gives no symmetry information about the core of the centre, because it has no response to symmetry-lowering stresses. However, in some cases the tightly bound particle can indicate a symmetry which is quite different from the symmetry of the centre in its relaxed ground state, as is shown in the second example.

4. The 4-Li "Q" centre – a hole trap

Luminescence with zero-phonon lines at 1044 and 1045 meV is observed in Li-doped silicon after radiation damage, Fig. 3 [12]. Isotope doping studies have shown that the centre contains 4 Li atoms [13]. First-principles calculations suggest that the ground state of the complex has T_d symmetry [14], but uniaxial stress measurements on the optical transitions indicate unambiguously a trigonal symmetry [15]. Relevant data are reproduced in Fig. 4. Before discussing them we note that the binding energy relative to the free exciton is 110 meV, and the centroid of the luminescence band is 100 meV below the zero-phonon lines – as for the ABC centre, the binding appears to originate mainly from the vibronic relaxation.

The vibronic relaxation may involve atomic motion which is totally symmetric in the original T_d point group, and it may involve deformations along a $\langle 111 \rangle$ axis to

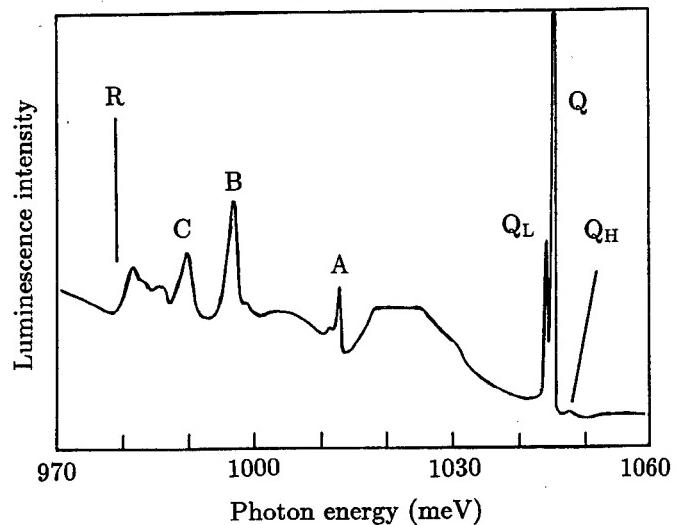


Fig. 3. Vibronic bandshape of the Li-related "Q" transition, recorded at 6 K, after Ref. [13]. The zero-phonon lines are labelled Q ("S = 0" line) Q_L ("S = 1" line) and Q_H for luminescence from the valley-orbit split-off electron state. The extent of the one-phonon sideband is shown by the sharp cutoff at the Raman phonon, indicated by 'R'. Peaks A, B, C are resonance modes produced by the Li atoms.

produce the observed trigonal symmetry. We can establish that it is predominantly trigonal. Zeeman measurements show that the lowest energy transition at 1044 meV is from an excited state which appears, at the spectral resolution, to be a spin triplet ($S = 1$) to a spin zero ground state, while the 1045 meV line is from an "S = 0" excited state to the

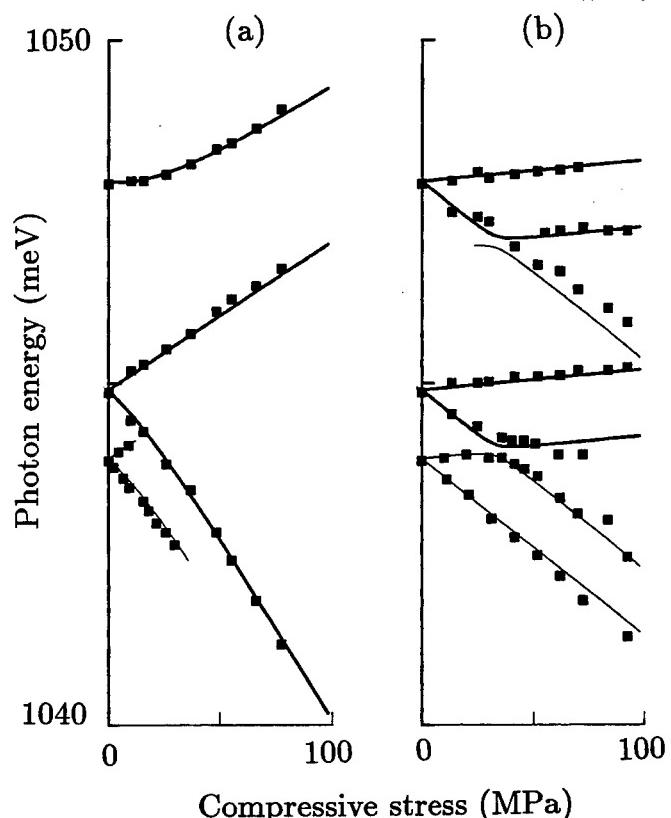


Fig. 4. The points show the effects of uniaxial stresses on the Li-related Q line (1045 meV), Q_L line (1044 meV) and the valley-orbit state Q_H (1048 meV) for (a) $\langle 001 \rangle$ and (b) $\langle 111 \rangle$ compressions [15]. The lines are calculated as described in Section 4. Thicker lines are for transitions originating (at low stresses) in the "S = 0" states, and thinner lines for "S = 1" states.

same spin zero ground state [16]. The spin states are immediate evidence that the orbital angular momentum of the hole has been quenched, that is, that the hole is in an axial field (with $\langle 111 \rangle$ orientation) at the centre [5]. To estimate the strength of the axial field we represent the valence band of silicon by the three degenerate orbital states P_x, P_y, P_z states; these states are mixed by the spin-orbit interaction to produce the $j = \frac{3}{2}$ and $\frac{1}{2}$ valence band maxima. For example, the $j = \frac{3}{2}, m_j = \frac{1}{2}$ state has the form $\sqrt{\frac{1}{6}}(P_x + iP_y)\downarrow - \sqrt{\frac{2}{3}}P_z\uparrow$ and the $j = \frac{1}{2}, m_j = \frac{1}{2}$ state is $\sqrt{\frac{1}{3}}(P_x + iP_y)\downarrow + \sqrt{\frac{2}{3}}P_z\uparrow$. We can recover a pure $P_z\uparrow$ state by taking the combination $\sqrt{\frac{3}{2}}(\frac{3}{2}, \frac{1}{2}) - \sqrt{\frac{1}{3}}(\frac{1}{2}, \frac{1}{2})$. For significant quenching of the orbital angular momentum, the axial field at the centre must be large compared to the spin-orbit interaction (44 meV for the valence band of Si). Using a perturbation V which is diagonal in the P_x, P_y, P_z set, so that $(P_x, VP_x) = (P_y, VP_y) = -Q, (P_z, VP_z) = 2Q$, we can calculate the mixture of the j, m_j states by this perturbation, and hence calculate the strength of the optical transitions from an electron in a totally symmetric orbital. The result is shown in Fig. 5, where allowance has been made for the different trigonal orientations in the crystal. The points are the observed relative strengths of the “ $S = 1$ ” to “ $S = 0$ ” transitions for a series of Li-related and S-related bands, with the perturbation taken to be equal to the observed binding energy, i.e. the energy difference of the free exciton and the zero-phonon line. For the 4-Li centre (labelled Q on Fig. 5) the axial perturbation required to produce the observed transition ratio of 1/70 [16] is closely equal to the total binding energy, which we have seen is equal to the relaxation energy – the atomic relaxation produced by binding the hole is essentially all axial.

We now have a picture for the bound exciton as consisting of a hole, which is tightly bound by about 100 meV in

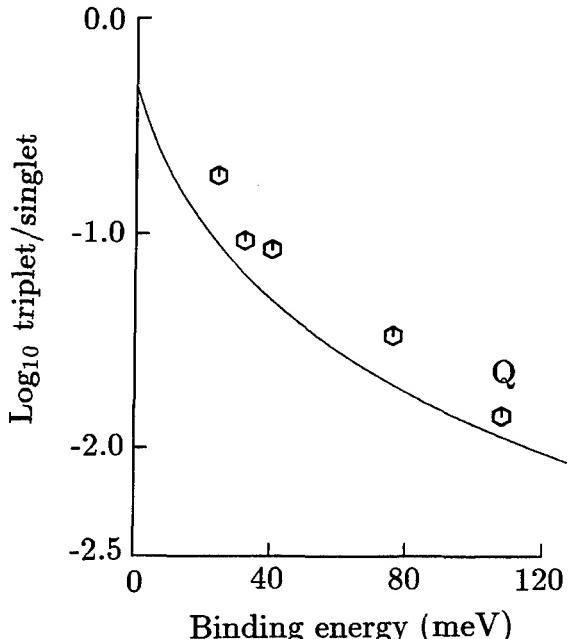


Fig. 5. Points show the ratios of the “ $S = 1$ ” and “ $S = 0$ ” zero-phonon lines for a series of similar Li-related vibronic bands in silicon [21], with the Li-related band discussed in section 4 labelled Q [20]. The exciton binding energy is the difference between the free-exciton energy and the observed zero-phonon line. The line shows the calculated ratio assuming a 44 meV valence-band spin-orbit coupling, and equating all the binding energy to the axial component.

the $\langle 111 \rangle$ axial field of the centre, plus an electron in relatively shallow states, bound by the Coulomb field of the trapped hole. As for the ABC centre, the 6 lowest conduction band states of the electron transform as A_1, E , and T_2 in the T_d symmetry of the unrelaxed centre. It turns out that the ordering is that the E state is lowest in energy and the A_1 state next. These states are not affected by the trigonal field of the centre (except for a possible shift in their energy centroid). However, a $\langle 001 \rangle$ compression produces a shift As from the hydrostatic component of the stress and additionally the E and A_1 electron states are perturbed according to the matrix [17]

$$\begin{matrix} & E_\theta & E_e & A_1 \\ E_\theta & -B_{el}s_\theta & B_{el}s_\theta & -\sqrt{2}B_{el}s_\theta \\ E_e & B_{el}s_e & B_{el}s_\theta & -\sqrt{2}B_{el}s_e \\ A_1 & -\sqrt{2}B_{el}s_\theta & -\sqrt{2}B_{el}s_e & E_a \end{matrix}. \quad (5)$$

Here the origin of the energy is the E state and the A_1 state lies E_a above it at zero stress. In terms of the stress tensors s_{ij} defined with respect to the crystal's cube axes,

$$s_\theta = 2s_{zz} - s_{xx} - s_{yy}, \quad s_e = \sqrt{3}(s_{xx} - s_{yy}). \quad (6)$$

Only one parameter, B_{el} , controls the splitting of the E state and its coupling to the A_1 state. The lines on Fig. 4(a) have been drawn with $B_{el} = 13$ meV/GPa, similar to the value of 11 meV/GPa for the conduction band states [11], confirming that the shallow-electron approach is correct, and rationalising the existence of the 1048 meV optical transition as the A_1 electron state. The hydrostatic term is $A = 4.25$ meV/GPa similar to the value for the exciton bound to the ABC centre, Section 3.

Under $\langle 111 \rangle$ stress, the electron states are not split, and all the splittings on Fig. 4(b) arise from the hole. Because the axial perturbation (~ 100 meV) considerably exceeds the spin-orbit coupling (of 44 meV) we can use a very simple description of the stress effects in which we ignore the spin-orbit interaction. The valence band is then represented by three degenerate orbitals, p_x, p_y, p_z where we use the conventional cubic axes x, y, z of the crystal. The non-degenerate hole state for the $[111]$ centre is the linear combination $(p_x + p_y + p_z)/\sqrt{3}$. An external compressive stress s along $[111]$ perturbs the centre by $As + 2Cs/3$ where the first term is for the hydrostatic component and C defines the coupling of p_i and p_j by a sheer stress s_{ij} as Cs_{ij} . Similarly the hole at a centre oriented along $[111]$ is expected to be perturbed by $As - 2Cs/9$. To fit the points on Fig. 4(b) requires $C = 32.6$ meV/GPa, considerably reduced from the valence band value of $C_v = 52$ meV/GPa [11].

To understand the reduction we introduce the language of the Jahn-Teller effect. A triply degenerate T_2 state (the p orbitals) in T_d symmetry can couple to a mode of vibration which transforms as T_2 . The coupling leaves a T_2 lowest in energy, but this is now a vibronic level, rather than a purely electronic level. An A_1 level from the first vibrational level at zero-coupling lies an energy 4β above the lowest level, where [18]

$$4\beta \sim 1.32E_{JT} \exp(-1.24E_{JT}/\hbar\omega). \quad (7)$$

In the limit of strong coupling, when the relaxation energy $E_{JT} \rightarrow \infty$, a four-fold degenerate ground state is obtained,

corresponding to distortions of the centre along the four $\langle 111 \rangle$ axes. For the Li-related centre, $E_{JT} \sim 110$ meV, and taking the effective vibrational mode to have a quantum $\hbar\omega = 25$ meV (as a rough mean of the one-phonon quanta in Fig. 3), $E_{JT}/\hbar\omega = 4.4$, and the strong-coupling limit has almost been reached. Because the electronic ground state is replaced by a vibronic ground state spread out over the four distortion axes, the effect of a given perturbation on the lowest vibronic level is reduced relative to its effect on the purely electronic states. The reduction factor for a shear stress is $k(T_2) = 2/3$ for strong coupling (and is correct to within a few percent for the finite coupling here, Ref. [19]). The effect of sheer stresses on the valence band, $C_v = 52$ meV/GPa, will therefore be reduced to $C = 34.7$ meV/GPa for the trapped hole, very close to the value derived in the last paragraph. Both the electron and hole behaviour can thus be understood with remarkable precision simply in terms of the deformation potentials of the band edges.

Evaluating eq. (7), the tunneling splitting $4\beta \sim 0.62$ meV. There are no data on this value. (It is not observed optically, and will be too close to the zero-phonon line to be readily detected in a stress experiment). Using this value as a guide, we can estimate the extent of the trapping in each distortion in terms of the time τ required to tunnel from one distortion to another:

$$\tau \sim \hbar/\beta \sim 4 \times 10^{-12} \text{ s.} \quad (8)$$

τ is in any event substantially smaller than the radiative decay time of $9.5 \mu\text{s}$ for the $S = 0$ transition [20], and the different distortions can communicate with each other. This allows us to observe an unusual phenomenon. Under $[111]$ stress a repulsion occurs between the $S = 1$ state of the $[111]$ centres and the $S = 0$ states of the other orientations (such as the $[\bar{1}\bar{1}\bar{1}]$ centres). If the trigonal axis of the centre was an intrinsic property, produced by its molecular structure, then this interaction would not occur, since the centres would be spatially separated in the crystal.

The situation here is that, in contrast to the usual cases considered in Jahn-Teller theory, there are two adiabatic surfaces, one for the " $S = 1$ " states and one for the " $S = 0$ " state, separated by $\Delta E = 1$ meV. Under $[111]$ stress, transitions from the $[111]$ centres decrease in energy and those from the $[\bar{1}\bar{1}\bar{1}]$ centres increase, bringing the two adiabatic surfaces into coincidence near $s = 9\Delta E/8C = 35$ MPa. At this stage a quantum-mechanical repulsion can occur between the lowest energy vibronic states associated with the two surfaces. If we were dealing with true $S = 1$ and $S = 0$ states there would be no observed repulsion, since the coupling is primarily through the orbital effects associated with the lattice strain and so would be zero between orthogonal spin states. However, because these are not pure spin states there will be a repulsion. The curves on Fig. 4(b) are calculated with a coupling of ± 2.5 meV/GPa between the different distortions, about an order of magnitude decreased from the first-order perturbation of the states. This reduction reflects the admixture of the valence band states in the " $S = 1$ " and " $S = 0$ " states. The problem is analogous to the ratio R of the optical transition intensities of the singlet and triplet lines, except that the transition intensity depends on the square of the wavefunctions. We expect the coupling to be reduced by $\sqrt{R} \sim 10$ from C , in agreement with the observed effect.

5. Summary

This paper has drawn attention to the role of lattice relaxations in binding excitons at isoelectronic centres in silicon. When an electron is tightly bound, an A_1 orbital is likely, and the consequent lack of orbital degeneracy implies that it is difficult to obtain data about the symmetry of the core of the centre. The hole moves in a relatively diffuse orbital, and so the symmetry observed for it may be the residual symmetry-lowering of the lattice observed at this large radius, or it may be a symmetry-lowering relaxation produced by the hole states. The symmetry of the relaxed ground state is difficult to obtain. This point has been emphasised for a tightly bound hole centre, where the trigonal symmetry has been shown to derive from lattice relaxation effects. However, it is then possible to use the well-established results of Jahn-Teller theory as developed for the limit of strong coupling to obtain a very precise understanding of the bound exciton states of these centres.

Acknowledgements

This work was supported by the Science and Engineering Research Council.

References

- See the review by Dean, P. J. and Herbert, D. C., in: "Topic in Current Physics" (Edited by K. Cho) (Springer, Berlin 1979), vol. 14, p. 55.
- Discussions of the binding mechanisms are given by, e.g. Faulkner, R. A., Phys. Rev. **175**, 991 (1968), and are reviewed extensively in Ref. 1.
- Hopfield, J. J., Thomas, D. G. and Lynch, R. T., Phys. Rev. Lett. **17**, 312 (1966).
- Thewalt, M. L. W., Watkins, S. P., Ziemelis, U. O., Lightowers, E. C. and Henry, M. O., Solid State Commun. **44**, 573 (1982); Labrie, D., Timusk, T. and Thewalt, M. L. W., Phys. Rev. Lett. **52**, 81 (1984).
- Gislason, H. P., Monemar, B., Dean, P. J. and Herbert, D. C., Physica **117B & 118B**, 269 (1983).
- Weber, J., Schmid, W. and Sauer, R., Phys. Rev. **B21**, 2401 (1980).
- Alt, H. Ch. and Tapfer, L., "Proc. of the 13th Int. Conf. on Defects in Semiconductors" (Edited by L. C. Kimerling and J. M. Parsey Jr.) (The Metallurgical Society of AIME, Pennsylvania 1985), p. 833; Tahjima, M., Japanese J. Appl. Phys. **21-1**, 113 (1981); Sauer, R., Weber, J. and Zulehner, W., Appl. Phys. Lett. **44**, 440 (1984).
- Pryce, M. H. L., in: "Phonons in Perfect Lattices and in Lattices with Point Imperfections" (Edited by R. W. H. Stephenson) (Oliver and Boyd, Edinburgh 1966), pp. 414–425.
- Zetsis, A. D., Chem. Phys. **40**, 345 (1979).
- Iqbal, M. Z., private communication (1993).
- Laude, L. D., Pollack, F. H. and Cardona, M., Phys. Rev. Lett. **B3**, 2623 (1971).
- Johnson, E. S., Compton, W. D., Noonan, J. R. and Streetman, B. C., J. Appl. Phys. **44**, 511 (1973).
- Canham, L. T., Davies, G. and Lightowers, E. C., J. Phys. **C13**, L757 (1980).
- Tarnow, E., J. Phys. Condensed Matter **4**, 1459 (1992).
- Davies, G., Canham, L. T. and Lightowers, E. C., J. Phys. **C17**, L173 (1984).
- Canham, L. T., Davies, G. and Lightowers, E. C., Inst. Phys. Con. Ser. **59**, 211 (1981).
- Wilson, D. K. and Feher, F., Phys. Rev. **124**, 1068 (1961).
- Caner, M. and Englman, R., J. Chem. Phys. **44**, 4054 (1966).
- Sakamoto, N., J. Phys. **C17**, 4791 (1984).
- Lightowers, E. C., Canham, L. T., Davies, G., Thewalt, M. L. W. and Watkins, S. P., Phys. Rev. **B29**, 4517 (1984).
- Lightowers, E. C. and Davies, G., Solid State Commun. **53**, 1055 (1985).

Lithium-Gold-Related Defect Complexes in n-Type Silicon

Einar Ö. Sveinbjörnsson

Department of Solid State Electronics, Chalmers University of Technology, S-412 96 Göteborg, Sweden

and

Sigurgeir Kristjánsson and Hafliði P. Gíslason

Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavík, Iceland

Received June 10, 1994; accepted June 15, 1994

Abstract

We demonstrate that lithium diffusion into gold doped n-type silicon at temperatures between 200 and 300°C results in the formation of two lithium-gold complexes. This was investigated using deep level transient spectroscopy (DLTS) combined with secondary ion mass spectroscopy (SIMS) and capacitance voltage (CV) profiling. One of the Au-Li complexes is electrically inactive and is observed indirectly as gold passivation. Virtually all passivated gold acceptors are reactivated after annealing for 30 minutes at 400°C and this process can be reversed by additional heat treatment at lower temperatures, as long as enough Li is still present in the crystal. This reaction can be described by a mass action law between negatively charged gold atoms and positively charged lithium ($\text{Au}^- + \text{Li}^+$) with a free binding energy of approximately 0.87 eV. The other Au-Li complex has a deep level, labelled L1, within the silicon band gap with an activation energy of 0.41 eV. The L1 signal is strongest after annealing at temperatures between 250 and 300°C, but the passivating complex appears to be favoured at lower temperatures. From the dissociation kinetics of L1 during reverse bias annealing we find that the complex consists of one gold atom and one or more lithium atoms.

1. Introduction

Lithium is a fast interstitial diffuser in silicon with high interstitial solubility [1, 2]. Interstitial Li acts as a shallow donor with a defect structure which has been studied in detail [3]. Also, lithium readily complexes with other foreign atoms within the silicon crystal, for example boron and oxygen [4, 5]. Recently, Höhne [6] detected a gold-related electron paramagnetic resonance (EPR) spectrum in gold-doped silicon co-doped with lithium. Further work by Alteheld *et al.* [7] using EPR and electron nuclear double resonance (ENDOR) techniques revealed two separate lithium-gold-related centres in crystals highly doped with Au and Li. These two centres were identified as an orthorhombic Au-Li pair and a trigonal Au-Li₃ complex.

The purpose of this work is to examine whether lithium is able to neutralize electrically the gold centre [8, 9] in a similar way as hydrogen does [10]. Due to the high solubility of lithium in silicon it is possible to obtain uniform Li concentrations between 10^{14} – 10^{17} atoms/cm³ using diffusion temperatures below 300°C. This is in contrast to hydrogenation of silicon where uniform hydrogen density in the 10^{15} cm⁻³ range is only achieved after annealing in hydrogen ambient at very high temperatures, typically 1200–1300°C [11, 12].

2. Experimental details

The starting material was 3 inch n-type phosphorus doped (100) oriented floating zone silicon wafers with resistivities

between 8.5 and 11 Ωcm. The details of the gold doping have been given elsewhere [13]. After etching and cleaning the Au-doped samples were covered with 99.9% pure Li in mineral oil emulsion. The samples were then heat treated in a diffusion furnace in open quartz boats using argon ambient. The usual procedure was a Li pre-deposition at 300–350°C for 10–30 minutes, after which excess Li was removed off the surfaces of the samples. Li was driven in using heat treatment between 250 and 350°C followed by rapid quenching in liquid nitrogen. Finally the samples were polished and etched and the front sides were evaporated with gold or aluminium to obtain Schottky diodes. GaAl alloy was used for ohmic contacts.

Two types of reference samples received the same heat treatments as the Au-Li co-doped specimens: (1) samples doped with gold but no Li and (2) samples only doped with lithium. The gold acceptor concentration in the gold doped reference samples did not change during the processing. Finally, no deep levels were found in the lithium doped reference specimens.

CV profiling was used to estimate the active shallow donor concentration in the samples. The active donor concentration is the sum of the concentrations of phosphorus donors, lithium donors and possible lithium-oxygen (Li-O⁺) shallow donors [5]. In addition, SIMS was used to estimate the total Li content. The SIMS data was calibrated against standard Li implantations. This normally gives an accuracy of ±20%. Concentration depth profiles of the deep traps were obtained with DLTS depth profiling [13]. The electron capture cross-sections of the deep levels were estimated using majority carrier pulse filling [14].

3. Results

3.1. An estimate of the lithium concentration

We found that the concentration of electrically active lithium from CV profiling was comparable to the total Li content estimated with SIMS. Figure 1 compares the Li content estimated with these two techniques. The SIMS analysis revealed that the lithium density was in most cases uniform throughout the samples. However, some out-diffusion closest to the surface was observed in samples with lithium concentrations above approximately 1×10^{15} cm⁻³. This out-diffusion occurs most likely when the Schottky diodes are formed since the Li concentrations did not change appreciably during storage at room temperature. In

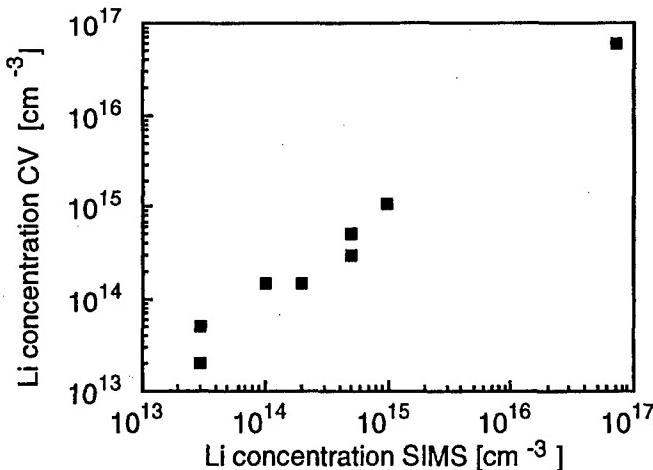


Fig. 1. Li concentrations in lithium doped n-type Si estimated using CV profiling and SIMS.

general, our observations were in agreement with the Li solubility studies of Pell [2].

SIMS analysis of reference samples not covered with lithium but heat treated in the same diffusion furnace typically showed Li contamination of approximately $1 \times 10^{13} \text{ cm}^{-3}$. In addition we searched for copper contamination both in Li doped and reference samples but no Cu was detected in any of the samples. The Cu detection limit was approximately $1 \times 10^{15} \text{ cm}^{-3}$. In addition DLTS revealed no known Cu-related traps [15, 16].

3.2. Neutralization of gold acceptors and the L1 peak

Figure 2 shows typical DLTS spectra of lithium diffused gold doped samples after heat treatments at different temperatures. After Li diffusion at 300°C the sample was slowly cooled to 100°C and kept at this temperature for 12 hours. This results in an active lithium donor concentration of $5 \times 10^{14} \text{ cm}^{-3}$. Thereafter the sample was polished and etched and divided into a number of $5 \times 5 \text{ mm}^2$ specimens, and each piece was then annealed at a specific temperature, as indicated in the figure, for 30 minutes followed by quen-

ching in liquid nitrogen. After surface cleaning, several Schottky diodes were made from each specimen.

The peak at 270 K corresponds to the gold acceptor level with activation energy $\Delta E \approx 0.54 \text{ eV}$ (including T^2 adjustment) and an electron capture cross-section $\sigma_n \approx 1 \times 10^{-16} \text{ cm}^2$ [13]. The peak labelled L1 has an activation energy of 0.41 eV and an electron capture cross-section of only $1.5 \times 10^{-17} \text{ cm}^2$. No temperature dependence of this capture cross-section was observed within the temperature range 225 – 256 K . This new signal was not observed in either type of reference samples. The small peak labelled G1 ($\Delta E = 0.19 \text{ eV}$, $\sigma_n = 1 \times 10^{-17} \text{ cm}^2$) has been studied previously [17] and arises from an acceptor level of a hydrogen-gold complex. G1 is only found in the surface region of the sample and is due to injection of hydrogen during wet chemical etching.

Reference samples doped with gold but no lithium show DLTS spectra that are similar to spectrum f, corresponding to gold concentration of approximately $2 \times 10^{14} \text{ cm}^{-3}$. We notice that gold acceptors are reactivated with increasing annealing temperature while the L1 peak is strongest at 300°C . In addition, we observed that the reactivation and passivation of gold acceptors is reversible; the amount of active gold acceptors is only dependent upon the final heat treatment, as long as no out diffusion of Li takes place. SIMS and CV analysis reveal that the lithium concentration within the samples normally remains constant during the annealing cycles. Occasionally, out-diffusion of Li was observed in samples where the lithium surface concentration had been reduced by cleaning and etching prior to the heat treatment. The passivation of gold is stable at room temperature; storage for several months has no effect on the DLTS spectra.

From a series of experiments using different Li concentrations we observed that the portion of gold which is neutral at the annealing temperature ($\leq 200^\circ\text{C}$) is not passivated by lithium while virtually all negatively charged gold atoms are passivated. So, in order to strongly passivate the gold atoms the Fermi-level must be well above the gold acceptor level at the annealing temperature.

The above results suggest that the L1 trap is related to both gold and lithium. This is even more apparent when the dissociation of L1 is studied during reverse bias annealing (RBA). Figure 3 shows the concentration depth profiles of

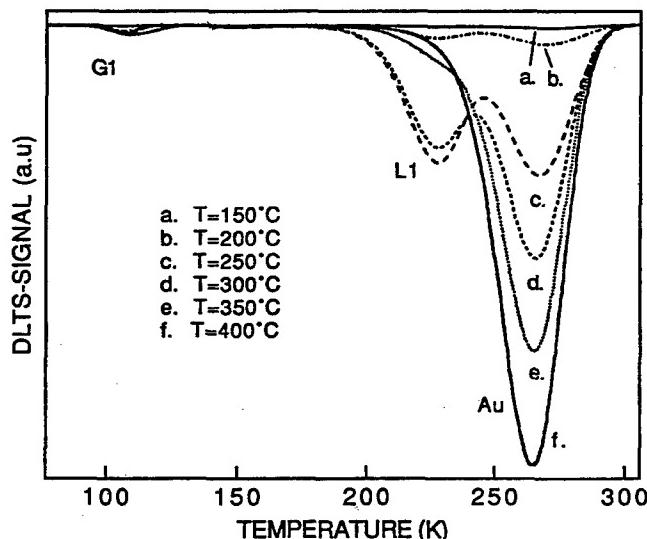


Fig. 2. Reactivation of gold acceptors as a function of temperature. The samples were first heat treated at 100°C for 12 hours and each sample was then annealed for 30 minutes at the specific temperature indicated. Gold acceptors are fully reactivated after annealing at 400°C for 30 minutes (sample (f)).

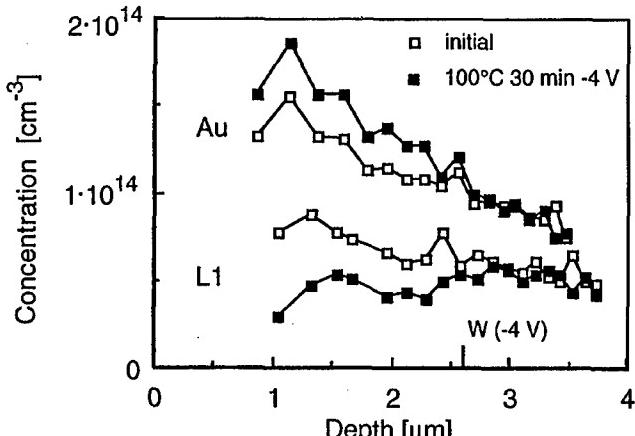


Fig. 3. Depth profiles of L1 and active gold acceptors before and after reverse bias annealing at 100°C for 30 minutes using 4 V reverse bias. The depletion layer edge, W, during the heat treatment is at approximately $2.6 \mu\text{m}$.

L1 and the Au acceptors before and after reverse bias annealing at 100 °C for 30 minutes. At this temperature the dissociation and the formation of L1 is in equilibrium, as is clearly evident outside the depletion region. However, the electric field within the space charge layer rapidly drives the positively charged lithium towards the surface, resulting in a net dissociation of L1 and an equal increase in the Au signal. This demonstrates the participation of gold in L1. Finally, Au-Li co-doped samples free from L1 traps do not show any change in the gold acceptor density after similar heat treatment.

4. Analysis and discussion

4.1. Au-Li pairs

The strong passivation of gold acceptors by lithium suggests that a long range Coulomb interaction between lithium donors and gold acceptors is responsible for the passivation. It is possible to describe the equilibrium concentration of such Au-Li pairs by the law of mass action:

$$[N_{\text{Au-Li}}^0]/[N_{\text{Au-}}][N_{\text{Li}^+}] = (1/N_{\text{Si}}) \exp [-\Delta G/kT] \quad (1)$$

where $N_{\text{Au-Li}}^0$ is the equilibrium concentration of neutral Au-Li pairs, $N_{\text{Au-}}$ is the fraction of the negatively charged gold given by Fermi-Dirac statistics, N_{Li^+} the lithium donor concentration, N_{Si} is the silicon lattice density, and ΔG is the Gibbs free binding energy of the pair. The concentration of Au-Li pairs is estimated by

$$N_{\text{Au-Li}}^0 = N_{\text{Au-t}} - N_{\text{Au}} - N_{\text{L1}} \quad (2)$$

i.e. we subtract both the remaining active gold density, N_{Au} , and the L1 concentration, N_{L1} , from the original gold concentration, $N_{\text{Au-t}}$. The following values were used for the parameters involved: $N_{\text{Au-t}} = 2 \times 10^{14} \text{ cm}^{-3}$, $N_{\text{Si}} = 5 \times 10^{22} \text{ cm}^{-3}$, and phosphorus donor density, $N_p = 5 \times 10^{14} \text{ cm}^{-3}$. The Au acceptor level is at $E_C - 0.54 \text{ eV}$. The lithium donor concentration was estimated with CV profiling and SIMS and was more or less independent of annealing temperature. We determined the binding energy of the Au-Li pair from an Arrhenius plot of equilibrium data, making use of the thermodynamic relationship at constant pressure and temperature:

$$\Delta G = \Delta H - T \cdot \Delta S \quad (3)$$

Such a plot is shown in Fig. 4. In all samples of Fig. 4 the equilibrium concentration of Au-Li pairs had been reached. The time needed to reach equilibrium was shorter than 30 minutes in the temperature range 150–400 °C. The free binding energy is defined as $E_b = -\Delta H$. A least squares fit to the data gives a binding energy of $0.87 \pm 0.04 \text{ eV}$ and a prefactor of $3 \times 10^{-23} \pm 2 \times 10^{-23} \text{ cm}^3$. The prefactor is close to $1/N_{\text{Si}}$ which suggests that the entropy term is small.

4.2. The L1 trap

The L1 trap is not stable at room temperature but is formed between 200–300 °C and can be frozen in by rapid quenching. Also, the L1 trap is absent at any temperature between 25 °C and 300 °C in samples highly doped with lithium ($[\text{Li}] \geq 5 \times 10^{16} \text{ cm}^{-3}$) where the Fermi-level remains well above the gold acceptor level. These observations suggest that the L1 trap belongs to a lithium-gold complex different from the Au-Li pair. The Au-Li pair is also clearly favoured energetically over the L1 trap at low temperatures.

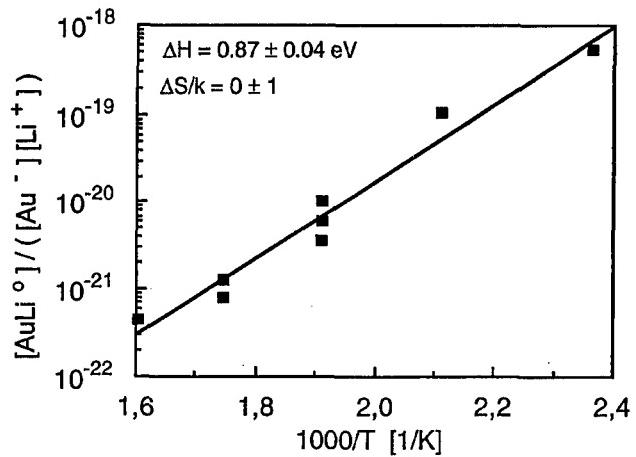


Fig. 4. Estimation of the binding enthalpy and entropy of Au-Li pairs from their equilibrium concentrations at various temperatures. A least squares fit gives a binding energy of $0.87 \pm 0.04 \text{ eV}$.

Alteheld *et al.* [7] reported two separate gold-lithium centres, identified as an orthorhombic Au-Li pair and a trigonal Au-Li_3 complex from EPR and Double ENDOR studies. The Au-Li pair in their work is most likely responsible for the gold passivation observed in the present work. Furthermore, the authors [7] observed that the Au-Li_3 complex has a single donor level within the band gap above $E_C - 0.4 \text{ eV}$. It is possible that the L1 trap in our work is the donor level of this complex, although the electron capture cross-section ($\sigma_n = 1.5 \times 10^{-17} \text{ cm}^2$) is smaller than expected for an attractive single donor trap.

4.3. Hydrogen vs. lithium passivation of gold

We find that both hydrogen and lithium are able to passivate the gold impurity in silicon but observe important differences especially regarding the charge state of the ions involved. In the case of lithium we observe two different complexes between gold and lithium where one of them is most likely the neutral Au-Li pair. The other complex has a deep level in the band gap labelled L1 and is possibly the previously reported Au-Li_3 complex [7]. Apparently negatively charged gold is also needed to form L1. The results indicate that opposite charge states of gold and lithium are needed for complex formation between the two ions. This differs markedly from hydrogen-gold complexes.

In the case of hydrogen we observe two complex centres between hydrogen and gold acceptors, one electrically inactive and other electrically active with a deep level in the band gap [18]. The electrically active centre is not formed between ions of opposite charge states while the inactive Au-H centre is most easily formed between Au^- and H^+ [18]. These complexes are stable at temperatures below approximately 150 °C. However, it is possible to transform these Au-H centres into one another by annealing and no free hydrogen is involved in these transformations. This is in contrast to the L1 gold-lithium complexes where depletion of free lithium donors results in a net dissociation of L1.

5. Conclusions

In summary, we observe two Au-Li complexes in n-type gold-doped silicon after in-diffusion of lithium. One of the complexes most likely consists of an Au-Li pair while the

other complex, L1, contains one gold atom and one or more lithium atoms. The proposed Au-Li pair has no deep levels in the upper half of the silicon band gap while the L1 complex has a deep level with an activation energy of 0.41 eV. The Au-Li pair is energetically favoured over the L1 complex at low annealing temperatures. Both complexes completely dissociate after 30 minutes at 400 °C and the gold acceptor concentration retrieves its original value.

Acknowledgements

We acknowledge Dr. Ulf Södervall (Chalmers University) for performing the SIMS analysis and Prof. Olof Engström (Chalmers University) for helpful discussions. This work was financially supported by the Swedish National Board for Industrial and Technical Development (NUTEK), Nordisk Forskerutdanningsakademi (NorFA) and the Icelandic Council of Science.

References

1. Pell, E. M., Phys. Rev. **119**, 1222 (1960).
2. Pell, E. M., J. Phys. Chem. Solids **3**, 77 (1957).
3. Watkins, G. D. and Ham, F. S., Phys. Rev. **B1**, 4071 (1970).
4. Reiss, H., Fuller, C. S. and Morin, F. J., Bell System Technical Journal **35**, 535 (1956).
5. Pell, E. M., J. Appl. Phys. **32**, 1048 (1961).
6. Höhne, M., Physica Status Solidi (b) **138**, 337 (1986).
7. Alteheld, P., Greulich-Weber, S., Spaeth, J.-M., Overhof, H. and Höhne, M., Materials Science Forum **143–147**, 1173 (1994).
8. Petersen, J. W. and Nielsen, J., Appl. Phys. Lett. **56**, 1122 (1990).
9. Watkins, G. D., Kleverman, M., Thilderkvist, A. and Grimmeiss, H. G., Phys. Rev. Lett. **67**, 1149 (1991).
10. Pearton, S. J. and Tavendale, A. J., Phys. Rev. **B26**, 7105 (1982).
11. Velorisoa, I. A., Stavola, M., Kozuch, D. M., Peale, R. E. and Watkins, G. D., Appl. Phys. Lett. **59**, 2121 (1991).
12. McQuaid, S. A. et al., Appl. Phys. Lett. **58**, 2933 (1991).
13. Sveinbjörnsson, E. Ö., Engström, O. and Södervall, U., J. Appl. Phys. **73**, 7311 (1993).
14. Lang, D. V., J. Appl. Phys. **45**, 3023 (1974).
15. Brotherton, S. D., Ayres, J. R., Gill, A., van Kesteren, H. W. and Greidanus, F. J. A. M., J. Appl. Phys. **62**, 1826 (1987).
16. Lemke, H., Physica Status Solidi (a) **95**, 665 (1986).
17. Sveinbjörnsson, E. Ö. and Engström, O., Appl. Phys. Lett. **61**, 2323 (1992).
18. Sveinbjörnsson, E. Ö. and Engström, O., Materials Science Forum **143–147**, 821 (1994).

Conversion Electron Mössbauer Spectroscopy Study of Iron Silicide Films Grown by MBE

M. Fanciulli and G. Weyer

Institute of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark

and

H. von Känel and N. Onda

Laboratory for Solid State Physics, ETH-Zürich, CH-8093 Zürich, Switzerland

Received May 4, 1994; accepted June 15, 1994

Abstract

Conversion Electron Mössbauer Spectroscopy has been applied to the study of different novel epitaxially stabilized phases of the Fe-Si system and also of Fe_3Si films. The silicides have been grown by Molecular Beam Epitaxy on Si(111). The ^{57}Fe Mössbauer parameters (isomer shift δ , linewidth Γ , quadrupole splitting Δ and magnetic field H at the nucleus) are reported and discussed in terms of the local surrounding of the Fe nucleus.

1. Introduction

The growth and the structural and electronic characterization of epitaxial iron-silicide phases on Si have attracted increasing interest during the last years from both the fundamental and technological points of view. In particular the semiconducting $\beta\text{-FeSi}_2$ phase, with a direct gap of 0.85 eV [1, 2] due to a Jahn-Teller-like instability, has potential optoelectronic applications when integrated with the well developed Si technology. However, recent theoretical [3] and experimental evidence [4] of a slightly lower indirect transition casts doubts on the applications of this compound for light emitting devices. The great versatility of Fe to form a number of epitaxial silicides offers a wide spectrum of potential applications and allows to address intriguing fundamental questions related to interface structure, growth kinetics and stability of epitaxial phases. Conversion electron Mössbauer spectroscopy (CEMS), due to the high sensitivity to the atomic and electronic surroundings of the Mössbauer nucleus, is a very powerful tool to investigate these questions and has been successfully utilized in the study of bulk phases of the Fe-Si system [5–8] as well as iron silicides [9].

Epitaxially stabilized ($\gamma\text{-FeSi}_2$, FeSi which has the CsCl structure and exists also in a $\text{Fe}_{1-x}v_x\text{Si}$ defective phase containing Fe vacancies v) and Fe_3Si phases have been grown [10–14] allowing the unambiguous identification of the Mössbauer spectra of the different metallic, semimetallic, semiconducting and magnetic silicides and their comparison with the known parameters of the bulk stable phases ($\epsilon\text{-FeSi}$, $\beta\text{-FeSi}_2$ and Fe_3Si). The Mössbauer parameters (isomer shift, quadrupole and magnetic splitting) for the different phases are reported and discussed with respect to the local surrounding of the Fe atoms.

CEMS is sensitive to layer thicknesses of up to $\sim 1500\text{ \AA}$. Samples for this study have been grown with natural Fe,

therefore only films thicker than $\sim 20\text{ \AA}$ have been investigated. The sensitivity can be increased to less than a monolayer as recently shown for silicides grown with enriched ^{57}Fe by MBE [9] or by laser ablation [15].

2. Experimental methods

The silicides were grown in a commercial MBE system by stoichiometric e-gun co-deposition of Fe and Si near RT on n-type Si(111) misoriented ($< 5^\circ$) substrates. Two Fe monolayers and a template of 10 \AA synthesized by codeposition of Fe and Si with FeSi or FeSi_2 stoichiometry represent the first steps in most growth processes. Structural and electronic properties were investigated by RHEED, XPS, STM, TEM, RBS and XRD. Details on the growth and characterization have been given elsewhere [10–14].

CEMS measurements have been performed at room temperature using a ^{57}Co source in a Rh matrix which was moved by a standard constant-acceleration drive. The samples were incorporated as electrodes in a parallel plate avalanche detector [16]. The isomer shifts are given relative to $\alpha\text{-Fe}$.

3. Experimental results and discussion

3.1. $\text{Fe}_{0.5}\text{Si}$

At the Fe/Si composition ratio of 1 : 2 two bulk phases are known [17]. The $\alpha\text{-FeSi}_2$, stable at temperatures between 967°C and 1223°C , is metallic and has a tetragonal lattice. The Mössbauer spectrum has been fitted by two quadrupole split lines [7] given in Table I. At temperatures lower than 967°C the semiconducting $\beta\text{-FeSi}_2$ is formed with an orthorhombic structure which results from a Jahn-Teller like instability of the fluorite phase [2]. Due to the orthorhombic structure this phase has two different Fe sites, both surrounded by a distorted cube of Si atoms, but with different Fe-Si distances [18]: Fe_I has $\text{Fe-Si} = 2.34\text{--}2.39\text{ \AA}$, while Fe_{II} has $\text{Fe-Si} = 2.33\text{--}2.44\text{ \AA}$. The distorted cube of Si atoms around the Fe sites produces an electric field gradient (EFG) at the nucleus which accounts for the observed quadrupole interaction [6, 7], while the observed isomer-shifts are determined by the average distances characterizing the two sites [6, see also Fig. 6 and related discussion].

Table I. Mössbauer parameters and relative Fe site fractions A_i of iron silicide phases (δ and H relative to α -Fe at R.T.).

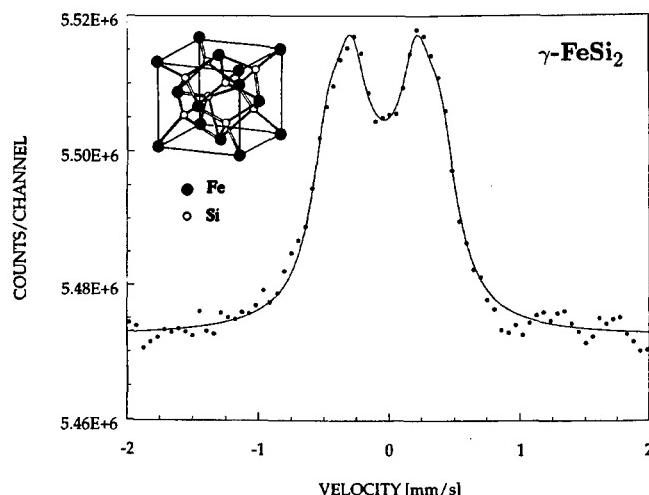
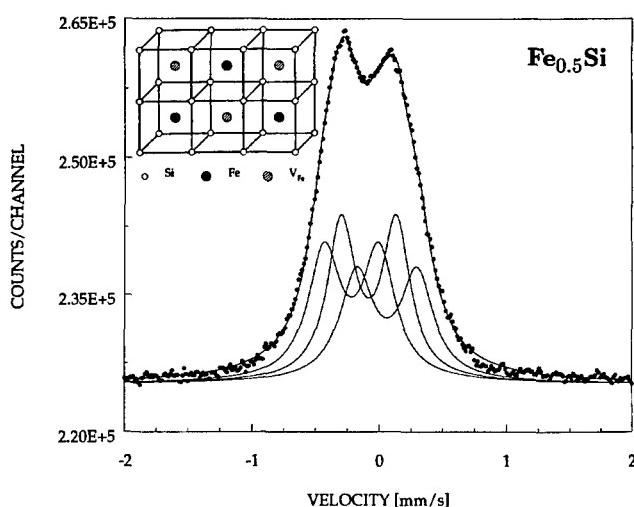
Sample	Thickness [Å]	Structure	Lattice parameters [Å]	Site	H/H_{Fe}	δ [mm/s]	Γ [mm/s]	Δ [mm/s]	$A_i/\sum A_i$ %
Fe_3Si	710	Do_3	$a = 5.66$	B [A, C] ₆ [A, C] ₅ [A, C] ₄ [A, C] ₃ Template and Interface Template and Interface	0.939(2) 0.868(2) 0.742(2) 0.599(2) 0.418(2) — —	0.09(1) 0.09(1) 0.18(1) 0.26(1) 0.34(1) 0.02(1) 0.56(1)	0.27(1) 0.49(1) 0.35(1) 0.28(1) 0.63(1) 0.23(1) 0.29(1)	— — — — — — —	20.1(5) 12.8(5) 7.3(5) 43.9(5) 9.5(5) 2.6(5) 3.8(5)
ϵ -FeSi [7]	Bulk	B20	$a = 4.46$		—	0.26(2)	0.34	0.51(2)	100
FeSi	890	CsCl	$2a = 5.54$		—	0.26(1)	0.24(1)	—	47.6(5)
—					—	0.25(1)	0.26(1)	0.26(1)	52.4(5)
α -FeSi ₂ [7]	Bulk	Tetrag.	$a = 2.69$		—	0.23	0.4	0.47	
—			$c = 5.13$		—	0.26	0.4	0.73	
β -FeSi ₂ [6, 7]	Bulk	Orthor.	$a = 9.863$	I	—	0.14	0.31	0.40	50
—			$b = 7.791$	II	—	0.03	0.31	0.44	50
—			$c = 7.833$						
γ -FeSi ₂	21	CaF ₂	$a = 5.387$		0.08(1)	0.15(1)	0.31(2)	—	100
—					—	0.19(1)	0.26(1)	0.43(1)	36.3(5)
—					—	0.05(1)	0.31(1)	0.47(1)	29.4(5)
—					—	0.32(1)	0.31(1)	0.43(1)	34.3(5)
$Fe_{0.5}Si$	740	CsCl	$2a = 5.40$						

Two epitaxially stabilized phases with a stoichiometry of 1 : 2 have been grown by MBE [10–14]. The bulk unstable fluorite γ -FeSi₂ phase [10–12] and the defective CsCl derived $Fe_{0.5}Si$ phase with Fe vacancies [12–14].

The γ -FeSi₂ phase has the fluorite structure and can be stabilized when the material is grown in the form of very thin films (21 Å in our case) due to the favourable lattice match to Si. No quadrupole splitting should be observed as the EFG at the Fe nucleus is zero due to the high symmetry at the Fe site surrounded by eight Si atoms at a distance $a_0\sqrt{3}/4$, 6 Fe vacancies at $a_0/2$ and 12 Fe at $a_0\sqrt{2}/2$. However, Christensen's finding [2] that the Fermi level is located in the strong and sharp peak of the Fe-d-DOS (density of states) indicates that γ -FeSi₂ could be magnetic. Indeed allowing for spin polarization, with the constraint of the fluorite structure, a ferromagnetic moment of $0.3\mu_B$ per formula unit develops [2] lowering the total energy of the highly unstable non-magnetic fluorite phase.

The CEMS for this phase, grown without template directly on the Si(111) substrate and annealed at 450 °C, is shown in Fig. 1 and the parameters used for the fitting are reported in Table I. The Mössbauer spectrum, fitted assuming a magnetically splitted sextet, yields $H/H_{Fe} = 0.08$ which is consistent with Christensen's ferromagnetic moment. This CEMS result represents a first experimental indication that FeSi₂ in the fluorite structure might be magnetically ordered. A fit assuming a quadrupole doublet gives a larger χ^2 and a quadrupole splitting much larger than the values observed in any other non-cubic silicide which makes this interpretation less likely. Furthermore in the case of CoSi₂, stable in the fluorite structure, only single lines have been reported in agreement with the high point symmetry at the Fe or Co nucleus [19].

The $Fe_{0.5}Si$ phase, which crystallizes in the CsCl structure and has 50% of Fe vacancies, shows three quadrupole split lines which can be associated to different Fe sites. The CEMS spectrum is shown in Fig. 2 and the fitting param-

Fig. 1. CEMS spectrum of γ -FeSi₂.Fig. 2. CEMS spectrum of $Fe_{0.5}Si$.

eters are reported in Table I. The first two doublets are consistent with a local lattice distortion towards the β -FeSi₂ phase. The intensity ratio of the doublet components is about 1 : 1, indicating an almost random distribution in the direction of the EFGs.

The fit of the Mössbauer spectrum yields a third doublet with an intensity ratio of 1 : 1 consistent with a random distribution of the Fe vacancies in the lattice. The larger isomer shift of this doublet is consistent with a local lattice distortion in which the Fe-Si distance is larger than the value in the ideal CsCl structure (from Fig. 6 about 2.39 Å as compared to 2.34 Å for the ideal structure).

3.2. FeSi

The bulk stable phase, at the composition ratio 1 : 1, is the ε -FeSi which has a cubic structure with four Fe atoms and four Si atoms in a unit cell. The local Fe symmetry is trigonal, with only one Si nearest neighbor at 2.29 Å, three Si second neighbors at 2.36 Å and three Si third neighbors at 2.53 Å [20]. The lower point symmetry at the Fe site results in the observed quadrupole split Mössbauer spectrum [7].

The Mössbauer spectrum of FeSi, epitaxially stabilized in the CsCl structure, is shown in Fig. 3 and the fitting parameters are reported in Table I. In this case there is only one Fe site with cubic point symmetry, therefore no quadrupole interaction is expected and a single line should be observed. However, beside the single line we have also found a doublet which is different from the reported value for the ε -FeSi phase. Epitaxially stabilized films should exhibit a coherent-to-incoherent transition at a thickness h_c before undergoing the first-order martensitic phase transition at the critical thickness H_c [21]. The MBE growth of FeSi yields coherent films in the CsCl structure characterized by an average trigonal distortion $\varepsilon_T \sim 3.3\%$ up to ~ 70 Å, above this critical thickness partial relaxation is observed [14]. The martensitic phase transition to ε -FeSi was found to be hindered kinetically for temperatures below 200 °C and thicknesses up to 1000 Å [14]. The observed quadrupole split line is therefore related to the fraction of the films characterized by a local trigonal distortion.

3.3. Fe₃Si

The ordered Fe₃Si has DO₃ structure and can be viewed as a fcc Bravais lattice with a basis consisting of four atoms, A,

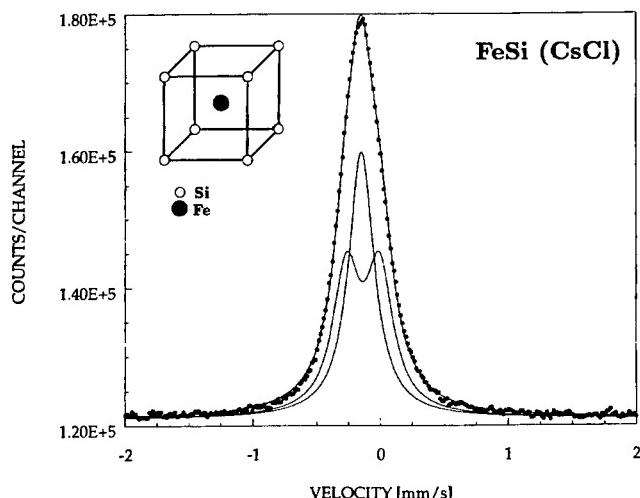


Fig. 3. CEMS spectrum of FeSi (CsCl).

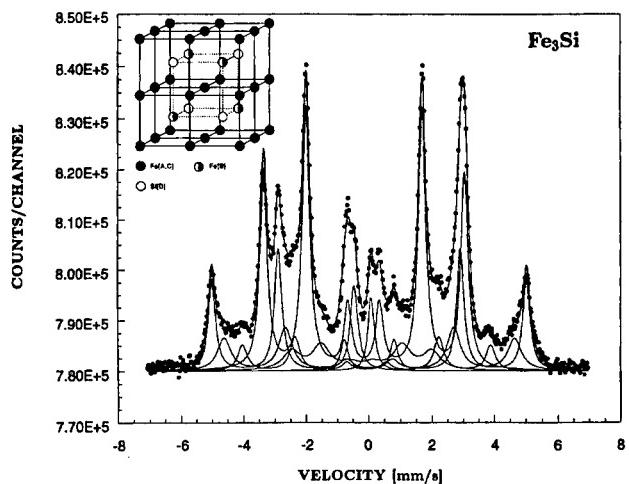


Fig. 4. CEMS spectrum of Fe₃Si.

B, C and D as shown in the inset of Fig. 4. There are two inequivalent (structurally and magnetically) Fe sites, those in the center of the cube (Fe_[B]) having cubic point symmetry, with 8 Fe_{[A, C]_n} nn and 6 Si nnn, and those on the corners (Fe_[A, C]) having tetrahedral point symmetry, with 4 Fe_[B] and 4 Si nn and 6 Fe_{[A, C]_{nnn}}. Polarized-neutron experiments [22, 23], saturation magnetization measurements [24] and Mössbauer spectroscopy [5, 8] on Fe-Si alloys have determined the magnetic moments on the two sites: $\mu_{\text{Fe}_{[B]}} = 2.2-2.4\mu_B$ and $\mu_{\text{Fe}_{[A, C]}} = 1.1-1.35\mu_B$.

The CEMS spectrum is shown in Fig. 4 and the fitting parameters are reported in Table I. The notation used in the text to identify [A, C] sites with n Fe nearest-neighbors is [A, C] _{n} . Since the fields expected in the case of Si excess ([A, C]₃) and Si deficiency ([A, C]₆, [A, C]₅) have been observed the presence in the films of non stoichiometric regions is inferred. However the relative intensities of the different fields are not consistent with a random distribution of the excess (deficiency) atoms in the films.

The results of Table I show a decrease of the field and an increase in the isomer shift for the [A, C] sites of $\partial H / \partial N_{\text{Si}} = -0.149$ and $\partial \delta / \partial N_{\text{Si}} = 0.08$ mm/s, respectively, as shown in Fig. 5, where N_{Si} is the number of silicon first neighbors.

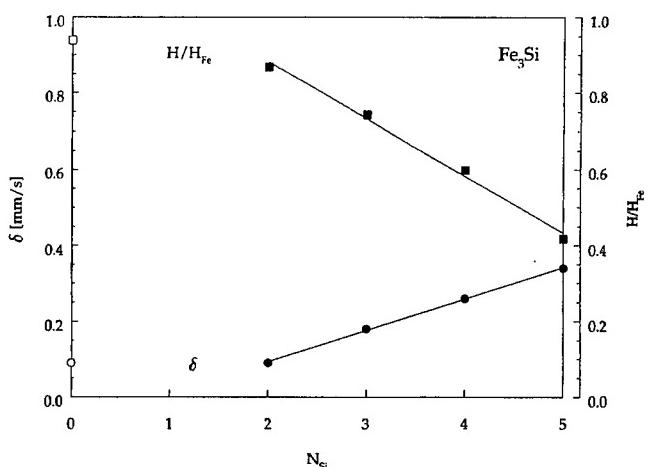


Fig. 5. Isomer shift and magnetic field as function of number of Si first neighbors for Fe₃Si.

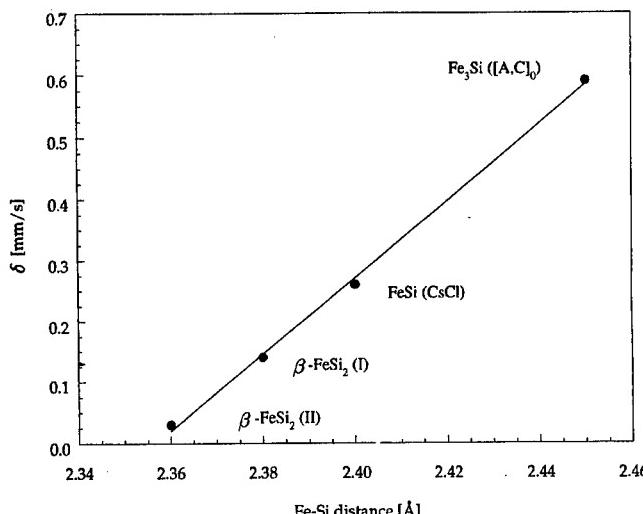


Fig. 6. Isomer shift as a function of Fe-Si distance for Fe sites with eight Si first neighbors in different compounds (the value for $\text{Fe}_3\text{Si} [\text{A}, \text{C}]_0$ was extrapolated from Fig. 5).

This is consistent with a reduction of $|\psi_{3s}(0)|^2$ due to different degrees of shielding of the 3s electrons by the d-like electron contributions of Si atoms [5, 25]. The isomer shift and the field at the B site are consistent with the six next nearest neighbors Si atoms.

4. Conclusions

The Mössbauer parameters for different novel epitaxially stabilized phases of the Fe-Si system as well as for FeSi_3 thin films grown by MBE have been reported. We found that for Fe sites characterized by eight first neighbors Si atoms the isomer shift scales linearly with the Fe-Si distance, as shown in Figure 6. An increase in the distance results in a reduction of the contact electron contribution and therefore in an increase of the isomer shift. As shown in

Figure 5 the isomer shift depends also on the number of Si atoms in the first coordination sphere of Fe [5, 6].

References

- Dimitriadis, C. A. *et al.*, *J. Appl. Phys.* **68**, 1726 (1990).
- Christensen, N. E., *Phys. Rev. B* **42**, 7148 (1990).
- Eppenga, R., *J. Appl. Phys.* **68**, 3027 (1990).
- Giannini, C., Lagomarsino, S., Scarinci, F. and Castrucci, P., *Phys. Rev. B* **45**, 8822 (1992).
- Stearns, M. B., *Phys. Rev. B* **29**, 1136 (1963).
- Wandji, R., Le Corre, C., Genin, J. M. and Roques, B., *Phys. Stat. Sol. (b)* **45**, K123 (1971).
- Helgason, Ö. and Sigfusson, T. I., *Hyp. Int.* **45**, 415 (1989).
- Rixecker, G., Schaaf, P. and Gonser, U., *Phys. Stat. Sol. (a)* **139**, 309 (1993).
- Degrote, S. *et al.*, *Mat. Res. Soc. Symp. Proc.* **320**, 133 (1993).
- von Känel, H., Stalder, R., Sirringhaus, H., Onda, N. and Henz, J., *Appl. Surf. Sci.* **53**, 196 (1991).
- Onda, N., Henz, J., Müller, E., Mäder, K. A. and von Känel, H., *Appl. Surf. Sci.* **56–58**, 421 (1992).
- von Känel, H., Mäder, K. A., Müller, E., Onda, N. and Sirringhaus, H., *Phys. Rev. B* **45**, 13807 (1992).
- Onda, N. *et al.*, *Appl. Surf. Sci.* **73**, 124 (1993).
- Onda, N. *et al.*, *Mater. Res. Soc. Symp. Proc.* **280**, 581 (1993).
- Fanciulli, M. *et al.*, "Fourth Seeheim Workshop on Mössbauer Spectroscopy", Seeheim, Germany, May 24–28 1994.
- Weyer, G., "Mössbauer Effect Methodology", **10**, 301 (1976).
- Kubaschewski, O., "Iron-Binary Phase Diagrams" (Springer Verlag 1982).
- Dusausoy, P. Y., Protas, J., Wandji, R. and Roques, B., *Acta Cryst. B* **27**, 1209 (1971).
- Vantomme, A., Wu, M. F., Dészi, I., Zhang, P. Q. and Langouche, G., *Hyp. Int.* **57**, 2133 (1990).
- Watanabe, H., Yamamoto H. and Ito, K., *J. Phys. Soc. Japan* **18**, 995 (1963).
- Bruinsma, R. and Zangwill, A., *J. Phys. (Paris)* **47**, 2055 (1986).
- Paoletti, A. and Passani, L., "Nuovo Cimento", **32**, 1449 (1964).
- Moss, J. and Brown, P. J., *J. Phys. F* **2**, 358 (1972).
- Hines, W. A. *et al.*, *Phys. Rev. B* **13**, 4060 (1976).
- Walker, L. R., Wertheim, G. K. and Jaccarino, V., *Phys. Rev. Lett.* **6**, 98 (1961).

Deep Centre Photoluminescence in Nitrogen Doped ZnSe

I. S. Hauksson, S. Y. Wang, J. Simpson*, K. A. Prior and B. C. Cavenett

Department of Physics, Heriot-Watt University, Edinburgh EH14 4AS, UK

and

W. Liu and B. J. Skromme

Department of Electrical Engineering and Center for Solid State Electronics Research, Arizona State University, Tempe, Arizona 85287-5706, U.S.A.

Received May 9, 1994; accepted June 14, 1994

Abstract

In this paper, we report optical studies of nitrogen doped ZnSe epilayers grown by molecular beam epitaxy. Photoluminescence spectra of the donor-acceptor pair region at different temperatures and different carrier concentrations show that two donors are present in the samples, residual shallow donors with activation energy 26 meV and deep donors with activation energy of 46 meV previously assigned to a V_{Se} -Zn-N_{Se} complex. In the exciton region we observe a new emission at 2.765 eV and the intensity increases when the epilayer is compensated by the deep donor. We therefore propose that this transition is related to a deep donor bound exciton. Excitation power dependent photoluminescence measurements show the presence of two deep transitions separated by 23 meV at low excitation intensities which we propose are due to deep donor-acceptor pairs.

1. Introduction

There has been a long history of attempts to dope ZnSe p-type. Park *et al.* [1] and Ohkawa *et al.* [2] were the first to successfully dope ZnSe grown by molecular beam epitaxy (MBE) using nitrogen as a dopant and this important step led to the first demonstration of II-VI blue-green semiconductor laser diodes [3, 4, 5]. Qiu *et al.* [6] first examined the photoluminescence spectra over a range of doping levels and showed that ZnSe:N is characterised by intense donor-acceptor pair (DAP) emissions. They reported p-type doping levels up to 10^{18} cm^{-3} but additional nitrogen is fully compensated and so we must understand the role of the nitrogen in the ZnSe in order to increase the number of active nitrogen acceptors.

We note that nitrogen forms an acceptor in ZnSe with activation energy of 110 meV. Detailed photoluminescence (PL) measurements on nitrogen doped ZnSe by Hauksson *et al.* [7] showed that a new DAP emission occurred in highly nitrogen doped ZnSe and it was shown that a deep compensating donor with a binding energy of 44 meV existed in more heavily doped material. It was proposed that the donor is a complex consisting of a nitrogen acceptor and a doubly charged selenium vacancy donor on a next nearest neighbour site (V_{Se} -Zn-N_{Se}). This complex would therefore be a single donor.

Recent Optically Detected Magnetic Resonance (ODMR) results [8] on the same samples showing both shallow and deep DAP luminescence, showed signals due to the shallow

isotropic donors and deep anisotropic donors consistent with our proposed model. Three ODMR resonance were observed; $g = 1.11$ for the 26 meV shallow donors, $g = 1.38$ for the 44 meV deep donor and $g = 2.00$ for the 110 meV deep nitrogen acceptor.

ZnSe : N samples were measured by C-V profiling four months after growth and we observed change in the carrier concentration [9]. Similar changes can also been seen in the photoluminescence spectra where the deep DAP transitions increase relatively to the shallow DAP over a period of six months. These results suggest that vacancies diffuse from the layer surface and the creation of deep donors therefore continues after growth at room temperatures. Calculations of vacancy concentrations as a function of temperature and selenium and zinc over pressures have shown that the material is always undersaturated with vacancies and that there is a substantial concentration gradient leading to the subsequent in diffusion of vacancies [9].

In this paper we examine the recombination process in nitrogen doped ZnSe grown by MBE and report detailed temperature dependence measurements of the emission for different nitrogen doping levels which show that the shallow and deep donors are associated with the nitrogen acceptors in the DAP recombination. Excitation power dependence measurements show the presence of deep impurity transitions possibly associated with deep donor-acceptor pairs.

2. Experimental details

The epilayers were grown by molecular beam epitaxy (MBE) using a Vacuum Generators VG288 growth system and details are given elsewhere [10]. For p-type doping we used a nitrogen rf plasma source supplied by Oxford Applied Research. For some samples part of the layer was illuminated by above bandgap light during growth using a Krypton ion laser (351 nm) with levels up to 4 W cm^{-2} . In the case of p-type dopant, the incorporation of active nitrogen is affected by the illumination and it has been shown that doping levels are increased [11]. For the samples concerned, a diameter of approximately 5 mm was illuminated through a heated quartz window. Thus a range of doping levels can be explored by observing the PL from different parts of the crystal. For example, the maximum nitrogen concentration will be at the centre of the illuminated region and the minimum outside the laser spot.

* Current address: Defence Research Agency, Great Malvern, Worcestershire WR14 3PS, UK.

Measurements were made on samples of various doping levels. The PL data presented here are from six samples, #328, #270, #276, #277, #278 and #279. All six samples are ZnSe layers of approximately 1.5 μm thickness on [100] GaAs substrate. Sample #328 was not illuminated during growth. The uncompensated acceptor concentration, $N_A - N_D$, where N_D is the donor concentration and N_A the acceptor concentration, was measured by electrochemical capacitance-voltage (*C-V*) profiling [12] and was found to be uniform in the direction of growth. Doping levels are given in Table I for each sample. The PL measurements were performed using cw Ar⁺ ion gas laser (351 nm). An Oxford cryostat was used to cool samples down to 4 K and the temperature could be varied up to room temperature. A SPEX 0.85 m double monochromator was used for the spectral analysis and the signal was detected with a cooled GaAs photomultiplier tube. In the measurements of the dependence of emission on excitation, the exciting light intensity was varied by using a variable neutral density filters.

3. Results

3.1. Bound excitons

Figure 1 shows the PL spectra in the exciton region from three regions of a nitrogen doped ZnSe at 4 K where the different regions correspond to unirradiated, partly irradiated and fully irradiated parts of the crystal and therefore correspond to different doping level. Spectra for all regions are dominated by acceptor bound exciton (A^0X) emission at 2.7893 eV which demonstrates that the sample is p-type. Free and donor bound exciton features can be seen for all regions. The lowest curve (a) in Fig. 1 shows the results from the unirradiated part of the layer. The emission at 2.758 eV is the phonon replica of the (A^0X) emission. The emission at 2.765 eV has not been discussed before as far as we know and the intensity increases with increasing nitrogen incorporation. We therefore conclude that this emission has its origin from the creation of a deep donor complex, such as excitons bound to the deep donors (D_{deep}^0X). This would correspond to an exciton binding energy of 38.1 meV compared to 10 meV for excitons bound to neutral acceptor. In curve (b) we see that the (D_{deep}^0X) emission has increased three times and then it is constant going from (b) to (c). The emission at 2.784 eV behaves in a similar way but we attribute this to excitons bound to pairs of acceptors separated by different distances in the lattice and this emission also increases as the acceptor concentration increases.

3.2. Donor-Acceptor pair emission

Figure 2 shows PL spectra at various temperatures for a sample which only shows shallow DAP emission ($N_A - N_D = 7 \times 10^{16} \text{ cm}^{-3}$). The spectrum at 4 K shows dominant

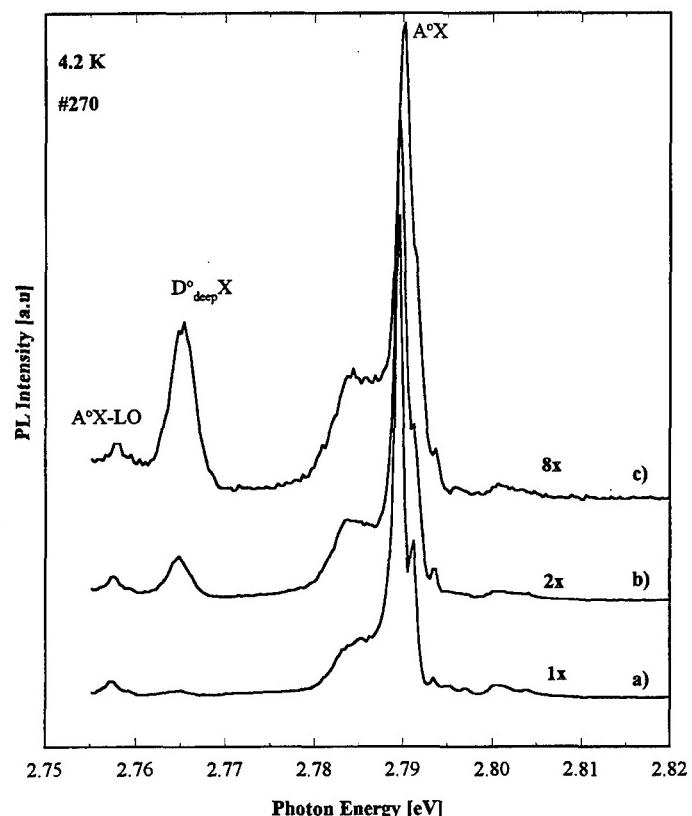


Fig. 1. PL spectrum for sample #270 taken at 4 K. The three different spectra are taken at different positions, (a) just outside the laser spot, (b) on edge of the laser spot and (c) just inside the laser spot. Going from a) to c) ($N_A - N_D$) changes from 1×10^{17} to $1.5 \times 10^{17} \text{ cm}^{-3}$.

shallow DAP emission and its phonon replica ($\hbar\omega_{\text{LO}} = 31.5 \text{ meV}$ for ZnSe). The free to acceptor (F-A) emission is observed as a shoulder on the high energy side of the no-phonon DAP emission. The Y₀ transition is also observed but this emission has been related to lattice defects. As we increase the temperature the DAP intensity and its phonon replicas decrease as the donors ionise and the (F-A) transition with phonon replicas becomes dominant at 50 K. At 145 K the donor-to-free hole (D-F) emission has become dominant indicating that the acceptors are partially ionised while at room temperature the D-F or free electron to free hole (F-F) emission dominates.

Figure 3 shows the PL spectra for a sample with more active nitrogen ($N_A - N_D = 2 \times 10^{17} \text{ cm}^{-3}$) and we observe two distinct DAP transitions. Strong evidence for the existence of two donors with different activation energies, as proposed by Hauksson *et al.* [7], is the difference in the quenching behaviour of the two different DAP transitions as temperature is increased. We clearly see that the shallow DAP emission decreases faster than the deeper DAP transitions and this demonstrates that there are two donors involved with different activation energies.

Table I. Active nitrogen concentration [cm^{-3}]

Sample no.	270	276	277	278	279	328
$N_A - N_D^*$	1×10^{17}	4×10^{17}	2×10^{17}	7×10^{16}	3×10^{16}	
$N_A - N_D^\dagger$	1.5×10^{17}	2.5×10^{17}	1.9×10^{17}	2.1×10^{17}	1.6×10^{17}	1×10^{17}

* Inside laser spot.

† Outside laser spot.

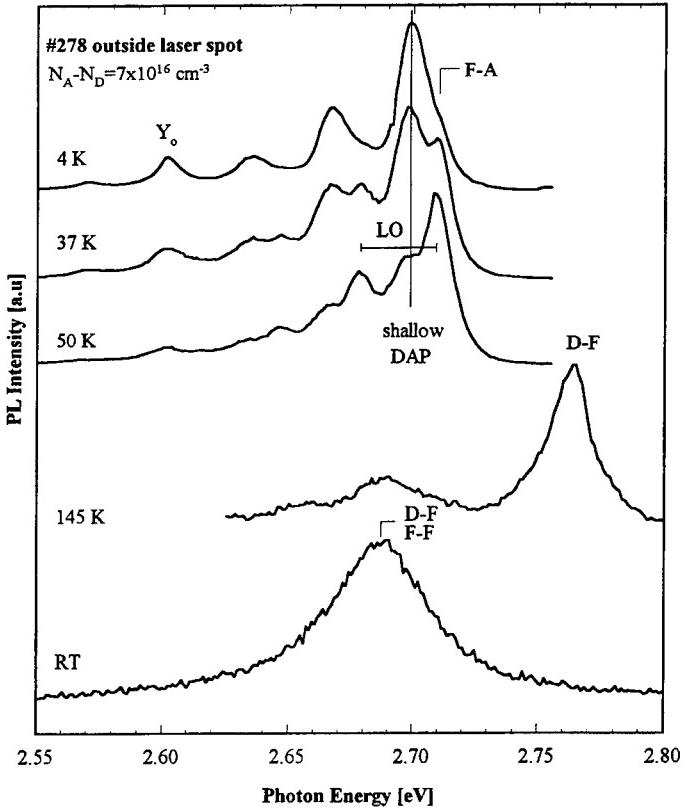


Fig. 2. PL spectrum of #278 at different temperature. The spectra are taken outside the laser spot where (N_A-N_D) is $7 \times 10^{16} \text{ cm}^{-3}$.

For temperatures higher than 40 K the intensity of the deep DAP emission decreases very rapidly, and at 70 K the F-A and deeper transitions dominate the spectra. In fact, these deeper transitions dominate the emission from highly doped material but even at this doping level they are clearly

observable. We will later show that these deeper transitions are DAP emissions which dominate the spectra for doping levels greater than $2 \times 10^{17} \text{ cm}^{-3}$. At higher temperatures, the luminescence evolves towards band-to-band recombination.

Figure 4 shows the PL spectra at different temperatures for more heavily doped sample than the one discussed above. Only deep DAP transitions and phonon replicas are observed at 4 K. When the temperature is increased a free-to-acceptor transition appears indicating that the 110 meV acceptor is still present. At higher temperatures (150 K), F-A, D-F and free-to-free (F-F) transitions dominate the spectrum and we can also see some features of deep transitions.

3.3. Deep emissions

Figure 5 shows the PL spectra for DAP emission under different excitation intensities. We have three sets of measurements for the irradiated regions of three different samples. Fig. 5(a) show PL from sample #278 which has an acceptor concentration of $2.1 \times 10^{17} \text{ cm}^{-3}$, Fig. 5(b) shows PL from sample #277 with an acceptor concentration of $2 \times 10^{17} \text{ cm}^{-3}$ and Fig. 5(c) show PL from sample #276 with an acceptor concentration of $4 \times 10^{17} \text{ cm}^{-3}$. The nitrogen flux during growth was lowest for sample #278 but highest for #276 so we expect less nitrogen incorporation for sample #278 and highest incorporation for #276. In Fig. 5(a), the PL is dominated by deep DAP transitions and we observe a shift of the no-phonon line to higher energy with higher excitation intensity. This is explained in terms of recombination between close DAPs compared with distant pairs. When we excite the samples harder, we saturate the

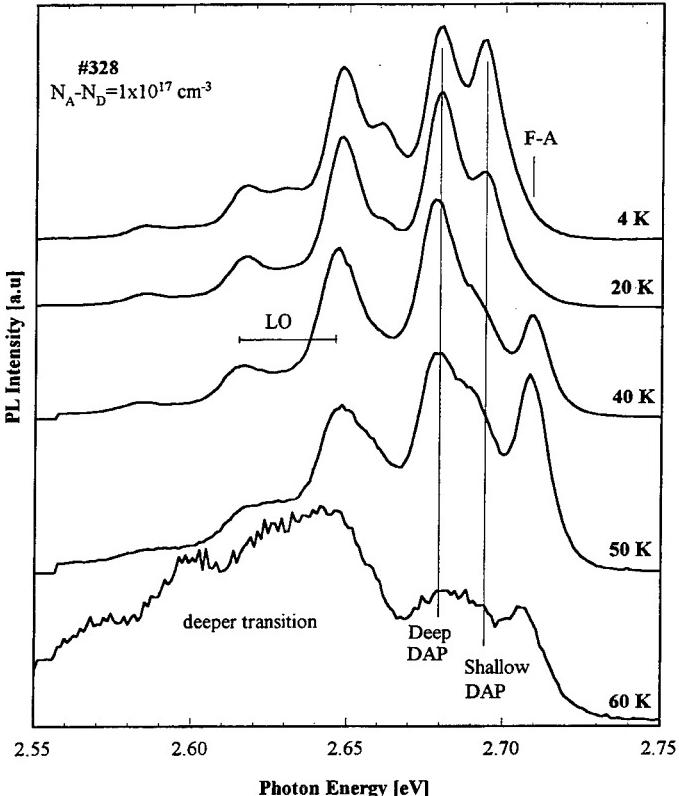


Fig. 3. Temperature dependent PL of #328. Excitation intensity is approximately 0.1 W/cm^{-2} .

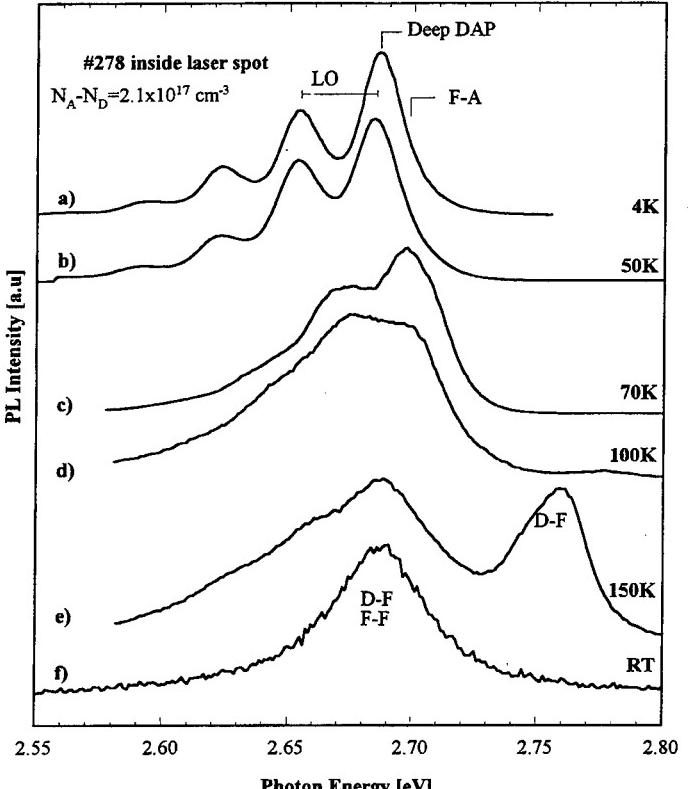


Fig. 4. Temperature dependent PL of sample #278 taken inside laser spot. Carrier concentration is (N_A-N_D) is $2.1 \times 10^{17} \text{ cm}^{-3}$.

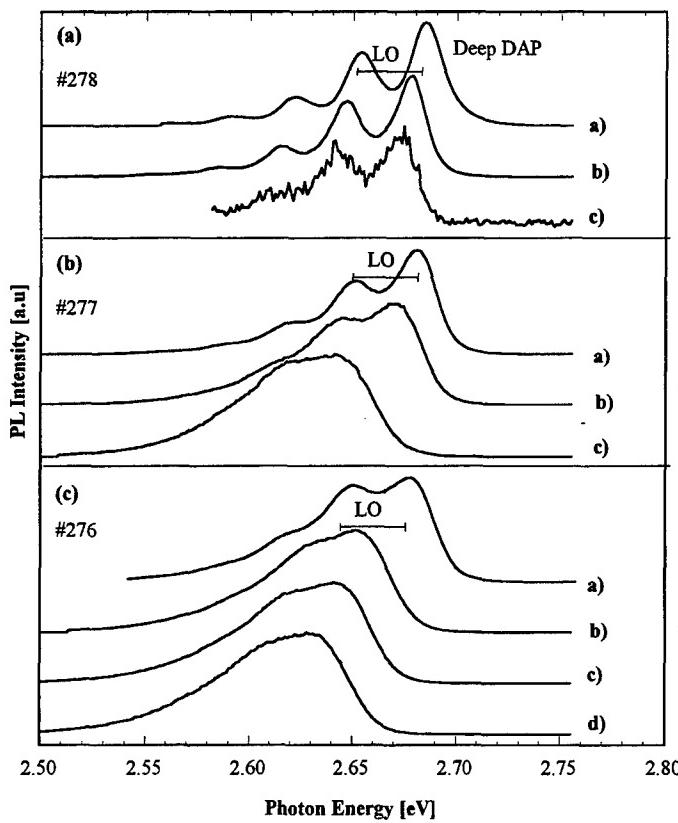


Fig. 5. PL spectrum taken at 4 K showing deep DAP transitions for different excitation intensity. Spectrum (a) to (c) show measurements from sample #278 inside laser spot and spectrum (d) to (f) show measurements from sample #277 inside laser spot. Excitation intensity is approximately: Fig. 5(a): (a): 3.5 W/cm^{-2} , (b): 0.1 W/cm^{-2} , (c): $2 \times 10^{-4} \text{ W/cm}^{-2}$, Fig. 5(b): (a): 3.5 W/cm^{-2} , (b): 0.35 W/cm^{-2} , (c): $3.5 \times 10^{-3} \text{ W/cm}^{-2}$ and Fig. 5(c): (a): 1.0 W/cm^{-2} , (b): 0.3 W/cm^{-2} , (c): 0.05 W/cm^{-2} and (d): 0.01 W/cm^{-2} .

recombination with larger r and recombination with small r increases, that is, the DAP emission shifts to higher energy. Going from $1 \times 10^{-4} \text{ W cm}^{-2}$ to 1 W cm^{-2} the peak shifts 11.8 meV compared to the shallow DAP emission which shifts 4.6 meV for the same intensity range. For low excitation intensity the Coulomb term is at a minimum since we are observing distant pairs and so we obtain the best estimation for E_D . Assuming that the Coulomb term is the same for shallow and deep DAP recombination we get a new estimation of 46 meV as a lower limit for the activation energy of the deep donor.

For sample #277 [Fig. 5(b)], curve (a) shows the PL spectrum for high excitation intensity and we can see the transitions which correspond to deep DAP emission with phonon replicas. From curve (b) to (c) the excitation intensity is decreased and the emission shifts to lower energy and the shape of the emission changes. For sample #276 [Fig. 5(c)], the highest excitation (curve (a)) shows emission which is also close to the deep DAP transitions but the peak on the lower energy side does not match the LO phonon

replica. As we lower the excitation intensity the spectrum is characterized by two peaks which are separated by approximately 23 meV. It is important to notice that the shape of this deep emission does not seem to change when the excitation power is changed.

4. Discussion and conclusions

In Fig. 5 we compare three samples with almost the same active nitrogen concentration ($N_A - N_D$) but we observe different PL spectra due to the different degree of incorporation of non-active nitrogen. Comparison by Qiu *et al.* [6] of the ($N_A - N_D$), measured by C-V, with the nitrogen concentration [N], as determined by secondary-ion-mass-spectroscopy (SIMS) shows that, typically ($N_A - N_D$) reaches the value $5 \times 10^{17} \text{ cm}^{-3}$ quite quickly but then saturates as compensation by donors occurs. A typical value for [N] for this saturation value is $1.8 \times 10^{18} \text{ cm}^{-3}$. The additional non-active nitrogen will form donor complexes which play a major role in changing the PL spectra. This kind of emission was first observed by Qiu *et al.* [6] and they proposed that this emission is related to DAP recombination where the donor level has broadened into an impurity band. However, the excitation PL measurements are not consistent with this model.

The PL data at different temperatures show that two different donors are present in ZnSe with doping concentration around $1 \times 10^{17} \text{ cm}^{-3}$, a shallow donor with activation energy of 26 meV and a deep donor with activation energy of 46 meV previously proposed as a $\text{V}_{\text{Se}}\text{-Zn-N}_{\text{Se}}$ complex. In highly doped ZnSe ($> 2 \times 10^{17} \text{ cm}^{-3}$) the excitation laser dependence of the PL shows an emission which is always characterised by two bands separated by 23 meV without LO phonon replicas. We propose that these transitions result from close (D_{deep}^0, A^0) pair recombination.

References

- Park, R. M., Troffer, M. B., Rauleau, DePuydt, J. M. and Haase, M. A., *Appl. Phys. Lett.* **57**, 2127 (1991).
- Ohkawa, K., Karasawa, T. and Mitsuyo, T., *J. Cryst. Growth* **111**, 797 (1991).
- Haase, M. A., Qiu, J., Depuydt, J. M. and Cheng, H., *Appl. Phys. Lett.* **59**, 1272 (1991).
- Jeon, H. *et al.*, *Appl. Phys. Lett.* **59**, 3619 (1991).
- Wang, S. Y. *et al.*, *Appl. Phys. Lett.* **61**, 506 (1992).
- Qiu, J., DePuydt, J. M., Cheng, H. and Haase, M. A., *Appl. Phys. Lett.* **59**, 2992 (1991).
- Hauksson, I. S., Simpson, J., Wang, S. Y., Prior, K. A. and Cavenett, B. C., *Appl. Phys. Lett.* **61**, 2208 (1992).
- Murdin, B. N. *et al.*, *Appl. Phys. Lett.* **63**, 2411 (1993).
- Prior, K. A. *et al.*, *J. Cryst. Growth* **138**, 94 (1994).
- Prior, K. A. *et al.*, *J. Cryst. Growth* **101**, 176 (1990).
- Simpson, J. *et al.*, *J. Cryst. Growth* **117**, 134 (1992).
- Wang, S. Y. *et al.*, *Appl. Phys. Lett.* **60**, 344 (1992).

Excitation of the 4f-Electron of Pr³⁺ in GaAs : Pr and Al_xGa_{1-x}As : Pr

P. L. Thee, Y. K. Yeo and R. L. Hengehold

Air Force Institute of Technology, Wright-Patterson AFB, OH 45433-7765, U.S.A.

and

G. S. Pomrenke

Air Force Office of Scientific Research, Bolling AFB, Washington, D.C. 20332-0001, U.S.A.

Received June 12, 1994; accepted June 15, 1994

Abstract

The luminescence intensity of Pr³⁺ varies dramatically with the Al mole fraction in Pr-implanted Al_xGa_{1-x}As. Two groups of major luminescence peaks have been observed near 1.6 and 1.3 μm, which can be attributed to the transitions of ³F₃ → ³H₄ and ¹G₄ → ³H₅ of Pr³⁺, respectively. For GaAs, the luminescence peak intensity near 1.3 μm is strong and the peak intensity near 1.6 μm is weak, whereas the opposite has generally been observed for Al_xGa_{1-x}As. Furthermore, only Al_{0.15}Ga_{0.85}As shows very strong luminescence peaks near 1.6 μm. This may be explained with a proposed excitation model for the 4f-electron of Pr³⁺ in Al_xGa_{1-x}As.

Rare earth element doped semiconductors are potentially valuable as light-emitting sources for various optoelectronic and photonic devices. High luminescence intensities of rare earths in III-V semiconductors, however, have not yet been achieved to a technologically practical level. This is partially due to the limited understanding of the excitation mechanism by which energy pumped into the host lattice is transferred to the rare earth 4f electrons. This situation is further complicated by the fact that the energy levels of rare earth-related deep centers, which are believed to be very important in 4f-electron excitations, are different for each rare earth element. For example, the Yb-related deep center in InP : Yb has an electron trap level at 30 meV below the conduction band [1], while Er-related deep centers in GaAs : Er have two hole traps at 35 and 340 meV above the valence band [2]. Among the rare earth elements, Er and Yb have been studied the most, but little work has been done on praseodymium (Pr) doped III-V semiconductors, even though Pr³⁺ is of considerable interest due to its emissions near 1.6 and 1.3 μm, which are respectively near the minimum attenuation and dispersion in silica-based optical fibers. Previously reported results of the photoluminescence (PL) of Pr³⁺ in III-V semiconductors include GaP : Pr [3–5], InP : Pr [6, 7] and GaAs : Pr [7–9]. Very recently, Erickson *et al.* [10] have reported PL spectra of GaAs : Pr along with calculations of crystal-field-split energy levels.

In order to better understand the excitation and de-excitation mechanisms of 4f electrons of Pr³⁺ in GaAs and Al_xGa_{1-x}As, and thereby to find the enhancing mechanism of the luminescence intensity, we have carried out a systematic study of Pr-implanted GaAs and Al_xGa_{1-x}As using low-temperature PL. This study has been performed as a function of bandgap energy with different Al mole fractions

and anneal temperatures, and an excitation mechanism for Pr³⁺ ions is suggested here.

PL was excited with an Ar-ion laser at a power of 200 mW, dispersed with a $\frac{3}{4}$ -m spectrometer using a 1.25 μm blazed grating, and detected with a liquid-nitrogen cooled Ge detector. Pr-ion implantation was carried out at room temperature at an energy of 390 keV with a dose of 5×10^{12} , 10^{13} , or $5 \times 10^{13}/\text{cm}^2$, which yielded a projected range of 0.09 to 0.1 μm depending on the Al mole fraction in Al_xGa_{1-x}As. Substrates used were LEC grown undoped GaAs, and MOCVD grown undoped Al_xGa_{1-x}As with x = 0.15, 0.3, and 0.5.

Figure 1 shows the PL spectra obtained from undoped GaAs and Al_{0.15}Ga_{0.85}As implanted with Pr and annealed at various temperatures. The emission spectra consist of two main PL peak groups: one group near 0.78 eV (1.59 μm), and the other group near 0.94 eV (1.32 μm). The former group consists of at least four main emission peaks at 0.755, 0.769, 0.779, and 0.783 eV, which are attributed to the intra-4f transitions between the crystal-field-split states of the excited level ³F₃ and the ground level ³H₄ of Pr³⁺. The latter group consists of at least five main emission peaks at 0.893, 0.898, 0.917, 0.931, and 0.945 eV, which are attributed to the intra-4f transitions between the states of the excited levels ¹G₄ and ³H₅ of Pr³⁺. Also, very weak transitions between ¹G₄ and ³H₄ have been observed at 1.15 eV and 1.20 eV, which were detectable only for the GaAs : Pr.

Figure 1 shows that for GaAs : Pr, the anneal temperature of 700 °C is certainly too low, and 850 °C is too high. Although the strongest emissions are seen at the 775 °C anneal, the emission intensities vary little over the range of 750 to 825 °C. Furthermore, the emission peak positions remain the same for the entire anneal temperature range. This suggests that the Pr³⁺ luminescence center is thermally very stable as would be expected for substitutional Pr on a Ga site. This observation, however, differs markedly from the anneal behavior of Er in GaAs, which exhibits a strong dependence of the 1.55 μm emissions on the anneal temperature [11, 12]. The anneal behavior of the Al_{0.15}Ga_{0.85}As : Pr shows that appreciable PL can be observed by annealing it even at 700 °C. The PL intensity increases with the anneal temperature up through 750 °C, exhibiting the best PL emissions for 775 °C with the signal decreasing above this temperature. Similarly to GaAs : Pr,

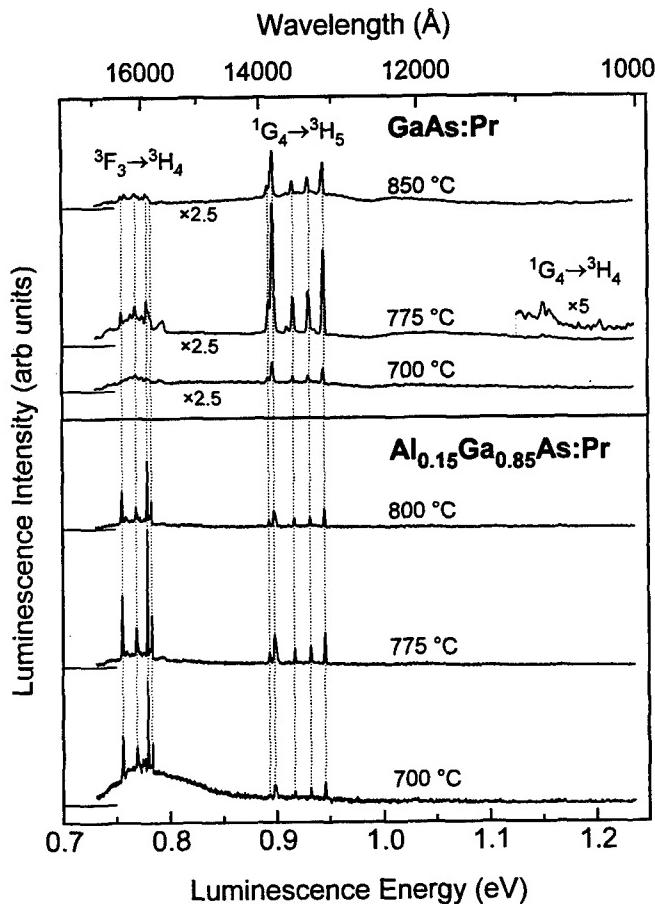


Fig. 1. Photoluminescence spectra taken at 3K for GaAs and $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$ implanted with Pr at 390keV with a dose of $10^{13}/\text{cm}^2$, and annealed at indicated temperatures for 15 sec using the rapid thermal annealing method. The zero PL level of each spectrum is indicated by a short horizontal line on the left axis.

the PL signal is very stable near 750 and 775 °C. The optimum anneal temperature obtained for both $x = 0.3$ and 0.5 hosts was 725 °C.

A PL study of Pr^{3+} was also conducted as a function of x for $\text{Al}_x\text{Ga}_{1-x}\text{As} : \text{Pr}$ and the results obtained from each sample with optimum anneal temperature and dose are shown in Fig. 2. It is clearly seen that each PL peak position in both 0.78 and 0.94 eV emission groups remains virtually the same for all samples. However, the PL intensities of the two luminescence groups vary dramatically as a function of x . For GaAs : Pr, the intensity of the 0.94 eV emission group is strong, but that of the 0.78 eV emission group is very weak. On the other hand, for $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As} : \text{Pr}$, the PL intensity of the 0.78 eV group increases drastically compared to that for GaAs : Pr, although the intensity of the 0.94 eV emission group remains about the same as that of GaAs : Pr. However, as x increases further to 0.3, the PL intensity of the 0.78 eV group now decreases considerably compared to that for $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As} : \text{Pr}$, and the 0.94 eV emission group almost disappears. For $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As} : \text{Pr}$, the PL intensity of the 0.78 eV group decreases much further. This behavior is very different from that observed from Er-doped AlGaAs [2, 13], which showed that the PL intensity of the Er rose with x through $x = 0.5$. This suggests that the excitation mechanisms for Er and Pr in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ may be different.

The excitation mechanism of the 4f electrons of Pr^{3+} in III-V semiconductors is essentially unknown. However,

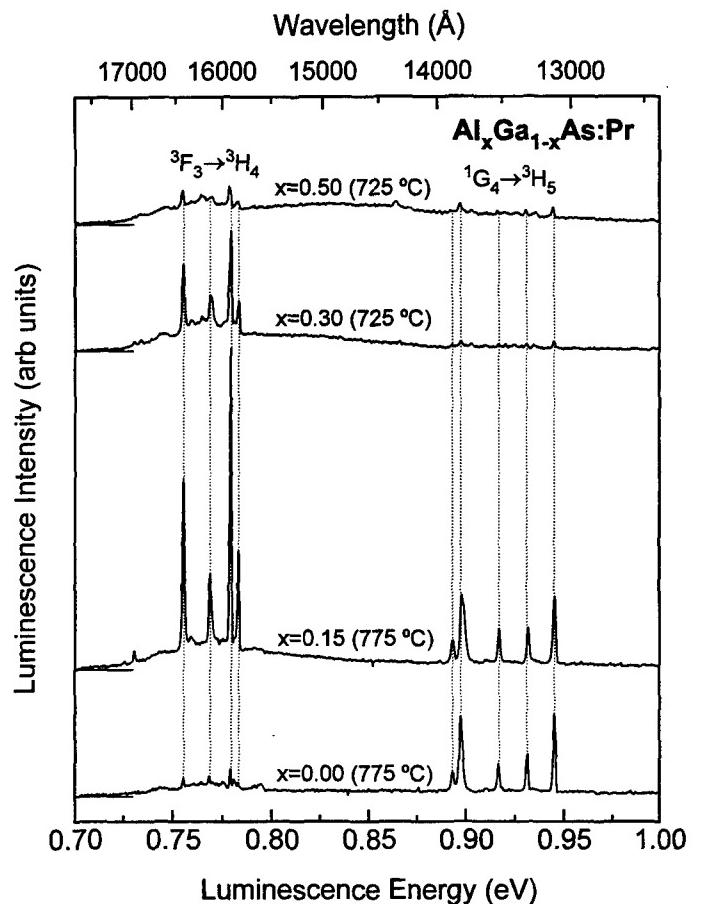


Fig. 2. Photoluminescence spectra taken at 3K for $\text{Al}_x\text{Ga}_{1-x}\text{As}$ with $x = 0, 0.15, 0.30$, and 0.50 implanted with Pr at 390keV with an optimum dose of 10^{13} , 5×10^{12} , 5×10^{13} , and $5 \times 10^{13}/\text{cm}^2$, respectively. Samples were rapid thermal annealed for 15 sec at indicated optimum temperatures.

present experimental results give insight into the possible excitation and de-excitation mechanisms of 4f electrons of the Pr^{3+} . In order to explain the excitation mechanism of the 4f electrons and the PL intensity dependence on Al mole fraction, it is assumed that the recombination energy of an exciton bound to the Pr^{3+} center is transferred non-radiatively to the 4f electrons. It is further assumed that the recombination energy could not only excite one 4f electron as is typically assumed for the rare earth in semiconductors, but could also simultaneously excite two 4f electrons, one from each of two closely located Pr^{3+} ions. This assumption has been made because the bound exciton recombination energy is generally too large to excite only one Pr^{3+} ion and to carry away the remaining energy through other particles and/or the lattice. At present, the Pr-related deep energy levels in $\text{Al}_x\text{Ga}_{1-x}\text{As}$ are unknown, as is the recombination energy of an exciton bound to Pr^{3+} . It is assumed that the trap level is somewhere around a few tens to hundreds of meV based on results for GaAs : Er and InP : Yb.

Figure 3 shows the center of gravity of each energy level manifold of Pr^{3+} in LaCl_3 [14]. These energy levels are used in the explanation of excitation and emission mechanisms for the currently observed PL results for Pr^{3+} . For GaAs : Pr, the bound exciton recombination energy is expected to be ~ 1.5 eV, and energetically, one 4f electron in the 3H_4 ground state can be excited to the 1G_4 level (~ 1.20 eV) with the remaining energy used to excite another 4f electron from the 3H_4 to the 3H_5 excited state (~ 0.26 eV).

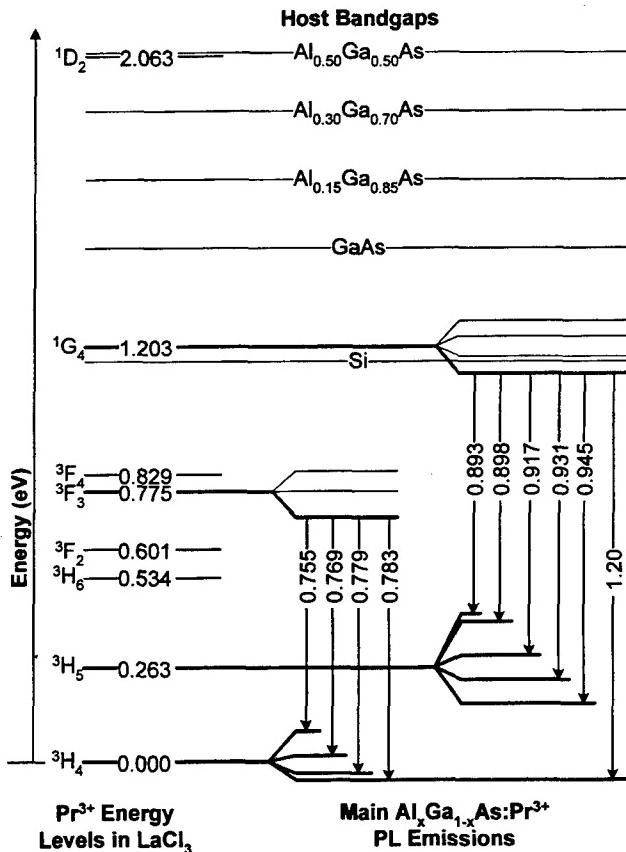


Fig. 3. Center of gravity of energy states of Pr^{3+} in an ionic crystal of LaCl_3 along with the host semiconductor bandgap positions. The observed main Pr^{3+} emission lines and associated crystal field splittings (exaggerated) are also shown. The numbers indicate energy differences between states in units of eV.

and/or carried away to the lattice in phonon form. Subsequently, the excited electron in the $^1\text{G}_4$ level can make a transition mainly to the $^3\text{H}_5$ level, emitting strong 0.94 eV group lines and make a very weak transition to the state $^3\text{H}_4$. Another possible excitation is the promotion of one $4f$ electron from the $^3\text{H}_4$ to either the $^3\text{F}_4$ (~0.83 eV) or $^3\text{F}_3$ (~0.78 eV) excited state with the simultaneous excitation of a second $4f$ electron from the $^3\text{H}_4$ to the $^3\text{F}_2$ excited state (~0.60 eV). However, no observation of the $^3\text{F}_2 \rightarrow ^3\text{H}_4$ emissions have been reported [10] from $\text{GaAs} : \text{Pr}$, implying that this route of simultaneous excitation of $4f$ electrons is at best comparatively weak. Further, this route is energetically less probable than the previous $^1\text{G}_4$ excitation route, thus the 0.78 eV group emissions are weak.

On the other hand, the case for $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As} : \text{Pr}$ is energetically very different from that for $\text{GaAs} : \text{Pr}$. The bound exciton recombination energy is expected to be ~1.7 eV. This energy is well matched to excite two $4f$ electrons from the ground states to either two $^3\text{F}_4$ excited states (~0.83 eV \times 2 = ~1.66 eV), two $^3\text{F}_3$ excited states (~0.78 eV \times 2 = ~1.56 eV), and/or one to the $^3\text{F}_4$ state and the other to the $^3\text{F}_3$ state. Note here that the crystal-field-split $^3\text{F}_4$ and $^3\text{F}_3$ states may be so close that the excited electron in $^3\text{F}_4$ could thermalize to the lowest level of the $^3\text{F}_3$ manifold. This is supported by the fact that no direct emissions from the $^3\text{F}_4$ -based transitions were observed in this study. At any rate, the electrons in the $^3\text{F}_3$ level, through the above excitation routes, can make transitions to the $^3\text{H}_4$ level, resulting in very strong 0.78 eV group line

emissions. Therefore, it is believed that the drastic increase in the intensity of the 0.78 eV group emissions of $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As} : \text{Pr}$ compared to those of $\text{GaAs} : \text{Pr}$ is due to the well-matched energy transfer. Also the bound exciton recombination energy could excite one $4f$ electron to the $^1\text{G}_4$ excited state (~1.20 eV) and the other $4f$ electron to the $^3\text{H}_5$ excited state (~0.26 eV), but energetically, this route would certainly not be as strong as the previous one. Therefore, the 0.94 eV group emissions remain strong, but not as strong as the 0.78 eV group emissions.

For $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As} : \text{Pr}$, the bound exciton recombination energy is expected to be ~1.9 eV. The excitation closest to this recombination energy would be that associated with the excitation of one $4f$ electron from the ground state to the $^1\text{G}_4$ excited state and the simultaneous excitation of the other $4f$ electron to the $^3\text{F}_2$ excited state. However, as mentioned above, the simultaneous excitation route of $4f$ electrons involving the $^3\text{F}_2$ is at best comparatively weak, and consequently emissions from the $^1\text{G}_4 \rightarrow ^3\text{H}_5$ transitions near 0.94 eV can also be weak, which is consistent with the present result. On the other hand, as for the case of $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$, the excitation of two $4f$ electrons from ground states to either two $^3\text{F}_4$ excited states, two $^3\text{F}_3$ excited states, and/or one to the $^3\text{F}_4$ state and the other to the $^3\text{F}_3$ state is still possible. However, these transitions are much less energetically favorable than those for $\text{Al}_{0.15}\text{Ga}_{0.8}\text{As}$ because of the larger energy mismatch. Thus, the PL intensity for the 0.78 eV group emissions is still strong, but is much weaker than that of the $\text{Al}_{0.15}\text{Ga}_{0.85}\text{As}$. For the indirect bandgap $\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$, the bound exciton recombination energy is expected to be ~2.06 eV, and the excitations discussed for the $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As} : \text{Pr}$ may still be applied with a possible additional transition involving $^1\text{D}_2$. However, the excitations and emissions will be reduced considerably from those for $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ because of the larger energy mismatch.

In conclusion, the emission intensity of Pr^{3+} in $\text{Al}_x\text{Ga}_{1-x}\text{As} : \text{Pr}$ varies dramatically with the energy bandgap depending on the Al mole fraction. Two main emission groups have been observed near 1.6 and 1.3 μm , which can be attributed to the transitions $^3\text{F}_3 \rightarrow ^3\text{H}_4$ and $^1\text{G}_4 \rightarrow ^3\text{H}_5$ of Pr^{3+} , respectively. This observation may possibly be explained using a model involving simultaneous excitation of two $4f$ electrons from two closely located Pr^{3+} ions via a non-radiative energy transfer from Pr -bound exciton recombination. The excitations are strong when they are energetically well-matched, otherwise they are weak.

Acknowledgement

The authors wish to thank G. Smith for his very valuable technical assistance.

References

- Whitney, P., Uwai, K., Nakagome, H. and Takahei, K., *Appl. Phys. Lett.* **53**, 2074 (1988).
- Elsaesser, D. W. *et al.*, *Growth* **127**, 707 (1993).
- Kasatkin, V. A., Kesamanly, F. P. and Samorukov, B. E., *Sov. Phys. Semicond.* **15**, 352 (1981).
- Kasatkin, V. A., *Sov. Phys. Semicond.* **19**, 1174 (1985).
- Gippius, A. A. *et al.*, *Mater. Sci. Forum* **10-12**, 1195 (1986).

6. Solomon, J. S., Pomrenke, G. S., Hengehold, R. L. and Yeo, Y. K., Proc. Seventh Intern. Conf. Secondary Ion Mass Spectroscopy (SIMS VII), 571 (1990).
7. Pomrenke, G. S., Hengehold, R. L. and Yeo, Y. K., Eur. J. Solid State Inorg. Chem. **28**, 159 (1991).
8. Pomrenke, G. S., Hengehold, R. L. and Yeo, Y. K., Inst. Phys. Conf. Ser. No. **106**, 339 (1989).
9. Pomrenke, G. S., Yeo, Y. K. and Hengehold, R. L., Mat. Res. Soc. Symp. Proc. **216**, 415 (1991).
10. Erickson, L. E., Akano, U., Mitchel, I., Rowell, N. and Wang, A., J. Appl. Phys. **74**, 2347 (1993).
11. Pomrenke, G. S., Ennen, H. and Haydl, W., J. Appl. Phys. **59**, 601 (1986).
12. Elsaesser, D. W., Colon, J. E., Yeo, Y. K., Hengehold, R. L. and Pomrenke, G. S., Mat. Res. Soc. Symp. Proc. **301**, 251 (1993).
13. Evans, K. R. et al., J. Vac. Sci. Technol. **B10**, 870 (1992).
14. Dieke, G. H. "Spectra and Energy Levels of Rare Earth Ions in Crystals" (Interscience Publishers, New York 1968), p. 200.

Passivation of Shallow and Deep Levels by Lithium in GaAs

Tryggvi Egilsson, Baohua Yang and Haflidi P. Gislason

Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland

Received June 10, 1994; accepted June 15, 1994

Abstract

Similar to hydrogen the group-I element lithium passivates various shallow and deep levels in GaAs. In p-type GaAs lithium passivates the shallow acceptors Zn_{Ga} and Cd_{Ga} and the deep acceptor Cu_{Ga} . In n-type GaAs it passivates native deep donors such as EL2 and EL6. However, in contrast to hydrogen there is no evidence of the passivation of shallow donors by lithium in GaAs. The passivation of shallow acceptors in p-type GaAs is inferred from a simultaneous increase of the Hall hole mobility and decrease in free carrier concentration throughout the bulk of Li-diffused samples. The acceptors can be reactivated by thermal annealing. We attribute the passivation to the formation of neutral Li-Zn and Li-Cd complexes. The passivation of Cu_{Ga} is concluded from the disappearance of deep level transient spectroscopy (DLTS) signals due to the Cu_{Ga} acceptor levels at $E_V + 0.15$ and $E_V + 0.40$ eV when the material is diffused with Li. A similar disappearance is observed for the well known PL band at 1.36 eV also attributed to the Cu_{Ga} defect. Both DLTS and PL signals can be reactivated by thermal annealing. Passivation of the native deep donors EL2 and EL3 in n-type GaAs is concluded from the reduction of the relevant peak heights in DLTS spectra after Li-diffusion. The defects can be reactivated by thermal annealing.

1. Introduction

Hydrogen passivation of shallow and deep impurities is a well known phenomenon both in elemental and compound semiconductors [1–5]. The interaction of hydrogen with defects in Si and GaAs has great technological significance. Doping with other group-I elements, on the other hand, may be important in order to clarify the mechanism by which hydrogen passivates. However, little information is available in the literature on the susceptibility of shallow or deep levels to passivation by other elements than hydrogen in any semiconductor.

All common shallow level acceptors in Si can be passivated by reaction with atomic hydrogen. Also, hydrogen passivation of the usual shallow donors can be obtained, although the passivation effect is not as strong as for the acceptors. In Si the passivation of deep levels by hydrogen is well known, but the mechanism by which the neutralization occurs is not as well understood as in the case of the shallow impurities [6]. Passivation by other species such as the alkali metals that could give important information on the nature of the bonding between the deep level impurities and hydrogen is scarce. Recently, however, it has been established that lithium causes strong passivation of the deep Au centre in n-type Si [7]. As far as shallow level passivation is concerned lithium has been reported to reduce the number of electrically active acceptors in Si [6].

Much work on shallow impurity passivation in GaAs has been carried out and it is now well established that hydrogen passivates both shallow donors [2] and acceptors [5] in GaAs. As far as the relatively high concentration of native

deep levels in as-grown GaAs is concerned, hydrogen is known to passivate a number of these levels, the most notable defect being the double donor EL2 [8, 9]. Also, passivation of the deep Cu acceptor [10] by hydrogen has recently been reported [4]. The details of the mechanism by which hydrogen neutralizes deep defects is even less investigated in GaAs than in Si. Doping by other group-I impurities has mostly been limited to lithium in fundamental investigations [11].

Passivation of electrically active defects in semiconductors is characterized by a reduction of the carrier concentration and a concurrent increase of the carrier mobility. This effect is caused by the formation of neutral pairs that reduces both the carrier concentration and the number of ionized scattering centres. In the case of electrical compensation, on the other hand, the reduced number of charge carriers is always accompanied by a decrease of the mobility brought about by an increase of the number of oppositely charged scattering centres. The effects of passivation on the mobility are likely to be screened by other effects such as compensation or inhomogeneity unless the passivated defect controls the concentration of majority carriers in the material. Thus, passivation of deep level defects does not enhance the carrier mobility in our samples although its effect on the DLTS spectra is quite dramatic.

2. Experimental

For Li-diffusions we used open quartz ampoules in an Ar ambient with the samples immersed in a saturated Ga-As melt prepared from 6N Ga metal, GaAs and 99.9% Li metal. The amount of Li in the melt ranged from 0.05–0.3 wt% depending on the shallow doping level of the starting material. After diffusion the samples were cooled to room temperature in the melt. A layer about 50 μm was removed from the surface of the Li-diffused samples after which they were polished and chemically etched. In most cases reference samples were made under identical conditions without Li. Cu diffusions were made in the same way, but with typically 5 wt% of 5N Cu in the Ga-As melt instead of Li. For DLTS measurements Schottky diodes were made by evaporating 1000 Å thick Al dots of diameter 1 mm onto p-samples but similar Au dots onto n-type samples. Ohmic contacts were made on the backside of the samples by welding Zn- or Sn-coated gold wire to the surface of p- and n-type samples, respectively.

The Li concentration profiles were analyzed by secondary-ion mass spectroscopy (SIMS) using a CAMECA IMS 4f system with a Li-implanted reference sample of

known fluence for calibration. The carrier concentration and mobility were determined with Hall effect measurements using the van der Pauw technique. Photoluminescence measurements were made at 14 K with the 514.5 nm Ar laser line as an excitation source. The signal was detected via a double 0.85 m Spex 1404 monochromator using a cooled Ge detector. DLTS measurements were performed in a system based on a 1-MHz Boonton 72B capacitance bridge and a double gate boxcar averager. Thermal scans were carried out in a closed-cycle He cryostat equipped with a feedback temperature controller.

3. Li-passivation of shallow acceptors in p-type GaAs

In order to investigate the effects of lithium diffusion on shallow acceptors a number of horizontal Bridgman p-type GaAs samples with Zn or Cd concentration 10^{17} – 10^{18} cm^{-3} were Li-diffused using the procedure described in the previous section. The diffusion conditions were chosen to ensure a Li concentration comparable to that of the shallow acceptors and a homogeneous distribution through the bulk of the samples.

Figure 1 illustrates the mobility enhancement in the temperature range 40–300 K caused by the lithium diffusion of two different starting materials at 600 °C. In both cases the lithium concentration was determined with SIMS measurements to be close to the shallow acceptor concentration. In both cases the mobility increases in the entire temperature range whereas the carrier concentration at room temperature decreases more than two orders of magnitude. The largest increase of the mobility, sixfold the original value, is observed in the Zn-doped starting material slightly above the temperature range between 50 and 100 K where the

dominating carrier scattering mechanism is ionized impurity scattering. The mobility enhancement is less pronounced in the Cd-doped samples, but still significant. We interpret the increased hole mobility in terms of a reduction of the number of electrically active acceptors caused by the formation of neutral donor-acceptor complexes.

Figure 2 shows the temperature variation of the hole mobility in three samples made from the same Zn-doped starting material, $N_p = 4 \times 10^{17} \text{ cm}^{-3}$. Li diffusion at 400 °C increases the hole mobility at all temperatures and at the same time decreases the hole concentration. This indicates passivation of the Zn acceptors as mentioned above. The Li concentration was determined by SIMS measurements to be similar to the shallow acceptor concentration [12]. The hole mobility of the Li-doped sample after heat treatment in pure gallium metal at 400 °C for 8 hours equals that of the as-grown starting material within experimental accuracy at all temperatures as illustrated in Fig. 2. The hole concentration of the annealed sample is also found to be the same as that of the starting material and the lithium concentration of the sample is not detectable by SIMS. This is true for all Li-diffused p-type samples measured in the present investigation.

By comparing the effect of Li-diffusion at various temperatures on the mobility we observed that diffusion at 400 °C was most effective. For example, Li diffusion at 680 °C reduces both the hole mobility and the hole concentration suggesting an increase in the number of positively charged scattering centres compensating the acceptors. This is also manifested by the high Li concentration measured in samples diffused in this temperature range.

From a number of carefully controlled Li diffusions and heat treatments at different temperatures we correlate both hole mobility and carrier concentration with the Li concentration of the samples. If passivation is to dominate over

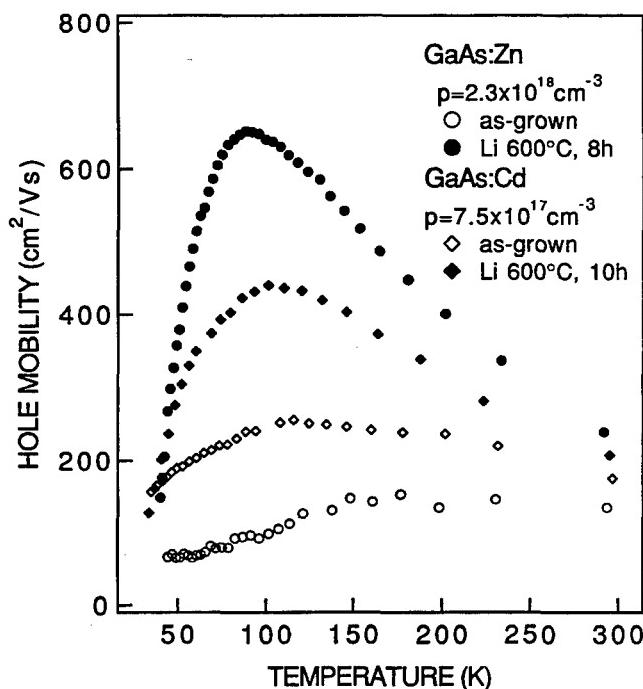


Fig. 1. Hall hole mobility as a function of temperature for as-grown GaAs : Zn and GaAs : Cd samples and two Li-diffused samples prepared as indicated in the figure. The Li concentration of the GaAs : Zn-Li sample is $1 \times 10^{18} \text{ cm}^{-3}$ and the hole concentration $N_{p300} = 3 \times 10^{16} \text{ cm}^{-3}$, but $N_{p300} = 2 \times 10^{15}$ for the GaAs : Cd-Li sample. The Li concentration of the latter was not measured.

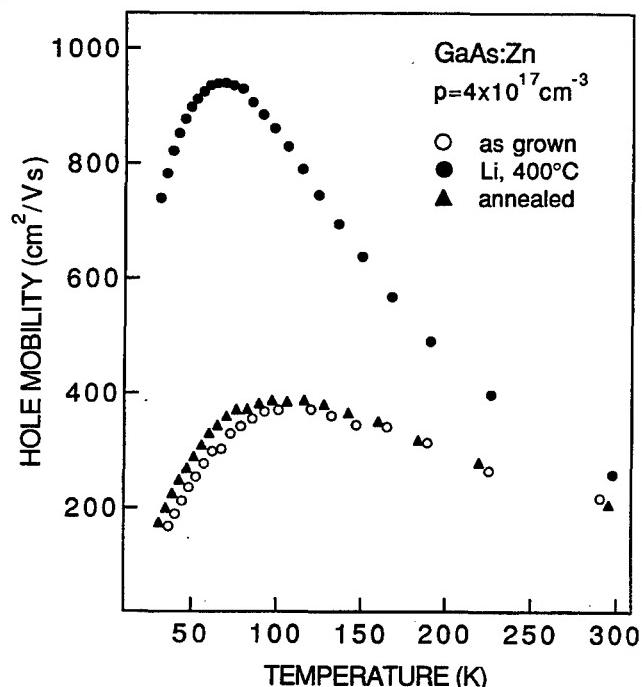


Fig. 2. Hall hole mobility as a function of temperature for (1) an as-grown GaAs : Zn sample, (2) a Li-diffused such sample and (3) a Li-diffused sample annealed at 400 °C for 8 hours in pure Ga-As melt. The heat treatment reactivates the Li-passivated Zn acceptors.

compensation in p-type samples the Li concentration must not significantly exceed the shallow acceptor concentration. At high Li concentrations compensation by randomly distributed Li-related donors obscures the passivation effect.

In Zn-doped GaAs hydrogen neutralizes Zn by bonding primarily to an As atom adjacent to the acceptor [5, 13]. The hydrogen is suggested to be stable in a bond-centred configuration between the Zn_{Ga} acceptor and the As atom. This position is not as natural for the larger Li atom when it passivates acceptors in GaAs, however. Cluster calculations predict lithium to have affinity to the interstitial site in Si [14] and the Li atom occupies an interstitial lattice site when it pairs with the B acceptor [15, 16] instead of the bond-centred Pankove configuration of hydrogen in the H-B pair [17, 18]. It seems natural to assume the same tendency for Li in GaAs.

Evidence from LVM spectroscopy of Li-Zn and Li-Cd pairs in GaAs in local vibrational mode measurements [19, 20] agrees with the passivation of Zn and Cd acceptors observed in this work. The vibrational spectrum of the Li-Cd pair in Li-diffused GaAs : Cd suggests that an interstitial Li donor passivates the substitutional acceptor by forming a neutral Cd-Li pair with a $\langle 001 \rangle$ axis [19]. This is the simplest configuration in the zincblende structure involving only one Li atom. In the case of Li-doped GaAs : Zn the vibrational spectrum is more complicated, which suggests that complexes of Li and Zn may involve more than one Li atom and have lower symmetry [20].

4. Compensation of shallow donors by lithium in n-type GaAs

The effects of lithium doping on shallow donors were studied in a wide range of undoped, Si- and Sn-doped n-type GaAs starting materials with different free electron concentrations. The samples were diffused to Li concentrations comparable to that of the shallow donors. Li diffusion of n-type GaAs always reduces the electron concentration in a similar way as found for the hole concentration of p-type samples. At a given diffusion temperature, however, the carrier concentration is more easily reduced in the p-type samples. In the present investigations we have not found evidence for passivation of donors. Instead, Li diffusion always causes the electron mobility to drop together with the electron concentration as expected for compensation. There is a further difference between Li-diffused n- and p-type GaAs. In a sharp contrast to the recovery demonstrated in Fig. 2 for Li-diffused p-type GaAs annealing at 400 °C has little effect on Li-diffused n-type samples.

The fact that local vibrational modes of Li paired with shallow donors can be observed in n-type starting materials [21] does not mean that the Li forms neutral pairs with the donors (which is our definition of passivation). Isolated Li_{Ga} is expected to be a double acceptor and if paired with a single donor it still acts as a single acceptor. Such Li-related acceptors would act as compensating centres that give rise to local vibrational modes of Li with donors. We conclude from our measurements that Li-related acceptors compensate the shallow donors in Li-diffused n-type samples. The observation that Li-diffused n-type starting material is resistant to annealing at 400 °C whereas Li-diffused p-type samples are restored by similar annealing agrees with the

above assumptions. Substitutional Li_{Ga} -related acceptors that compensate n-type materials can be expected to have higher dissociation energy than the $Li_i\text{-Zn}$ and $Li_i\text{-Cd}$ neutral complexes that passivate the Zn and Cd acceptors in p-type GaAs.

5. Li-passivation of deep copper acceptors in p-type GaAs

The passivation of deep Cu-acceptor levels by hydrogen and deuterium has been observed for the elemental semiconductors germanium [22] and silicon [23], as well as for the compound semiconductor gallium arsenide [4]. In all cases passivation was inferred from the reduction of relevant peak heights in DLTS spectra of samples that had been exposed to a hydrogen or deuterium plasma. We have observed that similar passivation can be brought about by Li diffusion.

Zinc-doped horizontal Bridgman GaAs with free hole concentration at room temperature $N_p = 2 \times 10^{16} \text{ cm}^{-3}$ was first diffused with Cu at 650 °C and then with Li at 400 °C. A number of samples were annealed at 400 °C to investigate the effects of post-diffusion heat treatment. Cu-diffused samples were heated for 4 hours at 400 °C in a saturated Ga-As melt, free of both Cu and Li. Cu-Li-diffused samples were heated for 6 hours at 400 °C in an Ar gas ambient. We summarize our DLTS and PL results in Figs 3 and 4.

Figure 3 shows DLTS curves for Cu- and Cu-Li-diffused samples. Curve a is the DLTS spectrum for GaAs : Zn after

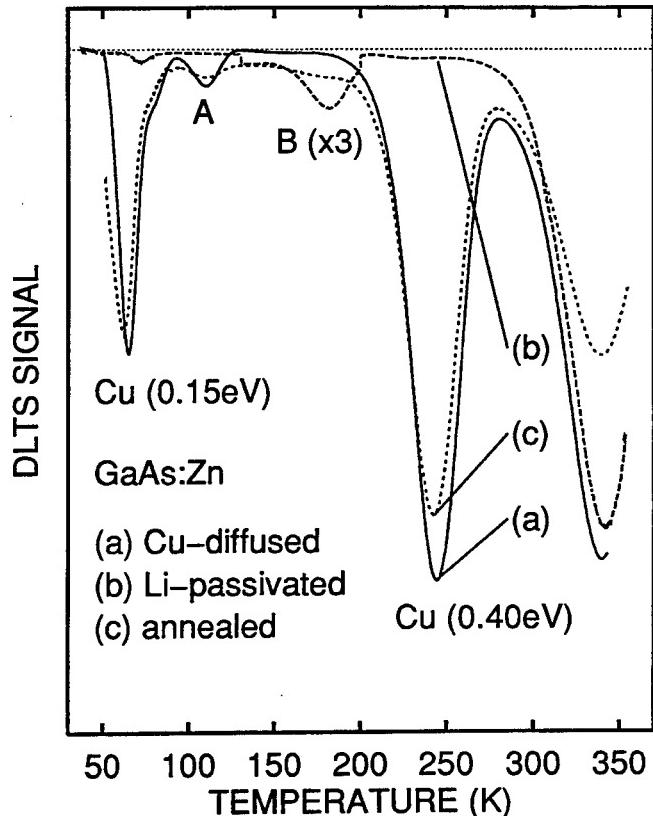


Fig. 3. DLTS spectra for GaAs : Zn after: (a) Cu diffusion for 5 hours at 650 °C; (b) Cu diffusion as in (a) and a subsequent Li diffusion for 4 hours at 400 °C. Both Cu-related peaks have disappeared; (c) annealing of the Cu-Li-diffused sample for 6 hours at 400 °C. The Cu-peaks have been reactivated. The measurement parameters for all curves are $V_R = 2 \text{ V}$, $V_p = 2 \text{ V}$ and $\tau = 0.414 \text{ ms}$.

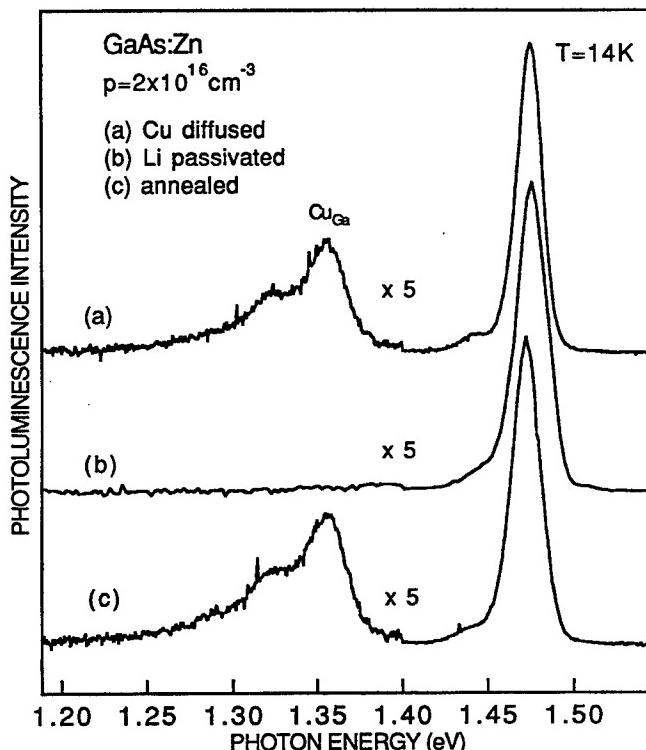


Fig. 4. Photoluminescence spectra for GaAs : Zn after: (a) Cu diffusion for 5 hours at 650 °C; (b) Cu diffusion as in (a) and a subsequent Li diffusion for 4 hours at 400 °C. Complete passivation of the Cu-related 1.36 eV PL band is observed; (c) annealing of the Cu-Li-diffused sample for 6 hours at 400 °C. The 1.36 eV band is reactivated.

Cu diffusion at 650 °C and, for reasons to be explained, an additional annealing for 4 hours at 400 °C in a saturated Ga-As melt. There are three main peaks in the spectra, at 65 K, 245 K and 340 K, of which the first two are caused by the copper diffusion [10]. Arrhenius plots were used to estimate the activation energy of the levels associated with the peaks. The peak at 340 K is due to a level with an activation energy of 0.58 eV and is also present in GaAs : Zn starting material. This level has not been investigated in detail, but its presence is presumably a result of iron contamination during crystal growth [24]. The peaks at 65 K and 245 K originate according to our DLTS measurements from levels with activation energies 0.15 and 0.40 eV, respectively. Using admittance spectroscopy it has been shown that 0.13 eV is a more accurate value for the activation energy of the shallower level [10]. The difficulty of determining its activation energy with standard DLTS measurements can be explained by carrier freezeout at low temperatures. Annealing of the Cu diffused material at 400 °C in a saturated Ga-As melt increases the amplitude of the DLTS peak at 65 K, but does not affect other parts of the spectrum. By monitoring the transient capacitance below 65 K we have seen that this may be attributed to a reduction in the peak quenching due to carrier freezeout. The increased peak height is therefore not due to a concentration increase of the responsible defect. As the Li diffusion and the annealing of the Cu-Li-diffused sample are carried out at 400 °C the DLTS curve for the annealed Cu-diffused sample (curve a) is the correct one to use in the comparison.

After Li diffusion at 400 °C both the Cu related DLTS peaks disappear from the spectra as can be seen from curve b in the figure. After annealing the Cu-Li-diffused material for 6 hours at 400 °C in an Ar gas ambient the Cu-related

DLTS-peaks reappear (curve c). A weak DLTS peak labelled A in Fig. 3 disappears with Li-diffusion but reappears with annealing. The identity of the peak is unknown but Li obviously affects it in a similar manner to the Cu-related peaks. There is another DLTS peak, B, visible on curve b in the figure, that appears after Li diffusion but disappears with annealing at 400 °C. The associated level has an activation energy 0.32 eV. In starting material and in solely Li-diffused material this level is not observed. We therefore suggest that a Li-Cu related defect is responsible for it.

Figure 4 shows three PL spectra. Curve a is the PL spectrum for GaAs : Zn after Cu diffusion at 650 °C. The main PL band at approximately 1.48 eV arises from transitions from donor or conduction-band states to shallow Zn acceptors. The Cu-related PL band at 1.36 eV reveals that Cu_{Ga} is present in the material [25, 26]. Also visible are the first two LO-phonon replicas of the 1.36 eV line. Curve b is the PL spectrum of the Cu-diffused material after a subsequent Li diffusion at 400 °C. The Li diffusion quenches the 1.36 eV luminescence but does not affect the 1.48 eV band. To examine whether the heat treatment by itself affects the PL spectra of Cu-diffused samples, some of them were annealed for 4 hours at 400 °C in a saturated Ga-As melt free from Li. This treatment has no effect on the PL spectrum. Heat treatment does, however, affect the Cu-Li diffused samples. The 1.36 eV luminescence band can be reactivated in Cu-Li-diffused samples by annealing for 6 hours at 400 °C in an Ar gas ambient. This corresponds to curve c in the figure.

Room temperature capacitance-voltage (CV) measurements on Schottky junctions were used to find the net acceptor concentration in the near-surface region of the samples. The Cu diffusion at 650 °C increases the concentration compared to the starting material. Annealing of the Cu-diffused samples at 400 °C in a saturated Ga-As melt does not change this. Li diffusion at 400 °C, on the other hand, reduces the concentration to a value slightly below that of the starting material. The Li diffusion therefore cancels the effect of copper and also gives rise to some extra concentration reduction. In light of our results concerning the passivation of Zn_{Ga} by Li we attribute the additional reduction to the formation of neutral Li-Zn complexes. Annealing of the Cu-Li-diffused material at 400 °C returns the net acceptor concentration to its value prior to Li diffusion.

From the DLTS results we conclude that lithium passivates the copper-related defects associated with the 0.15 eV and 0.40 eV acceptor levels. From PL measurements we conclude that a similar passivation occurs for the acceptor associated with the 1.36 eV pair luminescence. The coherent behavior of the Cu-related DLTS and PL signals agrees with the hypothesis that the signals are all due to the same defect, although it does not exclude other possibilities.

6. Passivation of deep donor levels in undoped n-type GaAs by lithium

Hydrogenation has been reported to passivate the main deep native donor levels in n-type GaAs [8, 9]. We have observed similar passivation due to Li-diffusion. In the investigation we used undoped horizontal Bridgman (HB) starting material with free electron concentration at room temperature $N_n = 1.5 \times 10^{16} \text{ cm}^{-3}$. Li was introduced into

samples by diffusion at 500 °C in a saturated Ga-As melt. For comparison we made a reference sample by annealing the as-grown starting material under similar conditions but without Li in the melt.

Figure 5 shows three DLTS spectra. Curve a is the spectrum for the starting material after annealing for 20 hours at 500 °C and serves as a reference. The annealing does not change the DLTS spectrum compared to the as-grown material. The spectrum is characteristic of n-type as-grown HB GaAs [27]. It has three main peaks at 160 K, 250 K and 350 K which we attribute to the native defects EL6, EL3 and EL2 respectively. Curve b is the spectrum after Li diffusion for 20 hours at 500 °C. The Li diffusion has a profound effect on the spectrum. The peak due to EL6 disappears and the EL2 peak is considerably reduced. Also, at least two peaks, at 200 K and 280 K, that are not observed in the starting material appear after the Li diffusion. We do not observe any new DLTS peaks in Li-diffused samples that represent levels of similar concentration as the EL6 and EL2 in the starting material. Curve c is the spectrum after the Li-diffused sample has been annealed for 20 hours at 500 °C in a saturated Ga-As melt free of Li. The spectrum has returned to its form prior to Li diffusion and shows a definite recovery of the EL6 and EL2 peaks. From room temperature CV measurements a decrease in effective donor concentration is observed in Li-diffused samples. Annealing at 500 °C returns the concentration to its value prior to Li-diffusion.

From the DLTS results we conclude that Li diffusion causes passivation of the deep donors EL6 and EL2 in GaAs. Further investigations are needed in order to estab-

lish the microscopic structure of the centres originating from the passivation.

7. Summary

We have observed that lithium can passivate both shallow and deep acceptor levels in p-type GaAs and deep donor levels in n-type GaAs. We see no evidence for the passivation of shallow donor levels, however. On the contrary, we have strong reason to believe that Li cannot passivate shallow donors in GaAs. The main experimental argument is that enhancement of electron mobility never accompanies reduction of the electron concentration by Li diffusion. In fact the mobility always decreases. Also the relatively high thermal stability of Li-diffused n-type GaAs contrary to observations in hydrogenated samples argues against passivation [6].

It has been shown that hydrogen has a negative charge state in n-type GaAs and that passivation of shallow donors is most likely due to Coulomb attraction between H^- and positively charged donors [28, 29]. Negatively charged Li has to our knowledge not been observed in any semiconductor and is in fact not expected to exist due to the low electron affinity of Li. As Coulomb interaction is commonly thought to be a strong driving force for the passivation of shallow levels, it is reasonable to assume that the charge state of the mobile lithium in n-type GaAs is a decisive factor in determining whether or not it passivates the shallow donors. As for the observed passivation by lithium of the deep donor EL2 in GaAs it is not likely to result from a simple Coulomb interaction because the Fermi level in the material at our diffusion temperature is well above the EL2 level which consequently is neutral.

Acknowledgements

This research was partially supported by the Icelandic Council of Science and the University Research Fund.

References

1. Pearton, S. J., Corbett, J. W. and Shi, T. S., *Appl. Phys.* **A43**, 153 (1987).
2. Chevallier, J., Clerjaud, B. and Pajot, B., "Semiconductors and Semimetals" (Edited by J. I. Pankove and N. M. Johnson) (Academic, Orlando 1990).
3. Jalil, A. et al., *Appl. Phys. Lett.* **57**, 2791 (1990).
4. Hofmann, G. et al., *Appl. Phys. Lett.* **61**, 2914 (1992).
5. Pajot, B., Jalil, A., Chevallier, J. and Azoulay, R. *Semicond. Sci. Tech.* **2**, 305 (1987).
6. Pearton, S. J., Corbett J. W. and Stavola, M., in "Hydrogen in Crystalline Semiconductors" (Springer-Verlag, Berlin Heidelberg 1992).
7. Sveinbjörnsson, E. Ö., Kristjánsson, S. and Gislason, H. P., these proceedings.
8. Lagowski, J., Kaminska, M., Parsey, Jr. J. M., Gatos, H. C. and Lichtensteiger, M., *Appl. Phys. Lett.* **41**, 1078 (1982).
9. Pearton, S. J., *J. Appl. Phys.* **53**, 4509 (1982).
10. Kullendorff, N., Jansson, L. and Lebedo, L. A., *J. Appl. Phys.* **54**, 3203 (1983).
11. Gislason, H. P., Yang, B. H. and Linnarsson, M., *Phys. Rev.* **B47**, 9418 (1993).
12. Yang, B. H., Gislason, H. P. and Linnarsson, M., *Phys. Rev.* **B48**, 12345 (1993).
13. Johnson, N. M., Burnham, R. D., Street, R. A. and Thornton, R. L., *Phys. Rev.* **B33**, 1102 (1986).

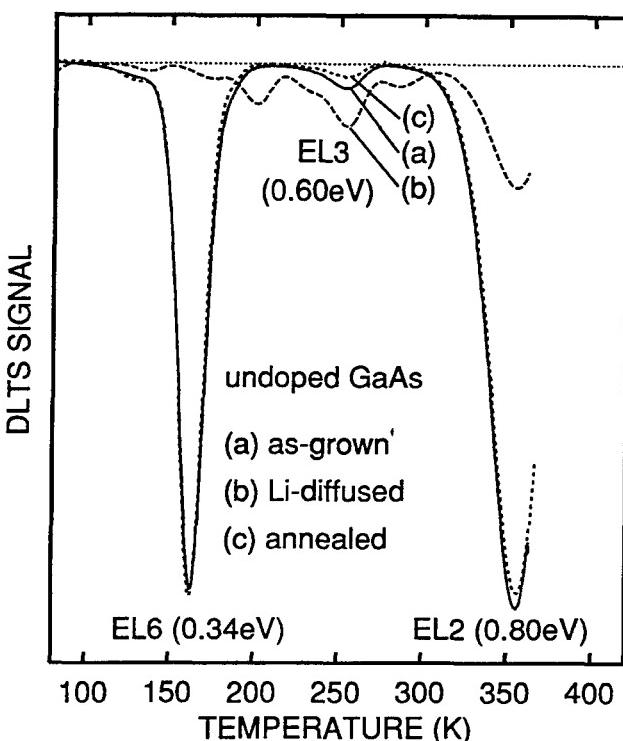


Fig. 5. DLTS spectra for undoped n-type GaAs after: (a) annealing for 20 hours at 500 °C in a saturated Ga-As melt free from Li; (b) Li diffusion for 20 hours at 500 °C. The main peaks in the spectrum, EL2 and EL6, have been considerably reduced: (c) annealing of the Li-diffused sample for 20 hours at 500 °C. The EL2 and EL6 peaks have been reactivated. The measurement parameters for all curves are $V_R = 2V$, $V_p = 2V$ and $\tau = 82.8$ ms.

14. DeLeo, G. G., Fowler, W. B. and Watkins, G. D., Phys. Rev. **B29**, 1819 (1984).
15. Balkanski, M., Elliott, R. J., Nazarewicz, W. and Pfeuty, P., in "Lattice Defects in Semiconductors" (Edited by R. R. Hasiguti) (Univ. of Tokyo Press, Tokyo 1968), p. 3.
16. DeLeo G. G. and Fowler, W. B., Phys. Rev. Lett. **56** 402 (1986).
17. Pankove, J. I., Carlson, D. E., Berkeyheiser, J. E. and Wance, R. O., Phys. Rev. Lett. **51**, 2224 (1983).
18. Pankove, J. I., Zanzucchi, P. J., Magee, C. W. and Lucovsky, G., Appl. Phys. Lett. **46**, 421 (1985).
19. Lorimore, O. G. and Spitzer, W. G., J. Appl. Phys. **38**, 3008 (1967).
20. Leung, P. C., Skolnik, L. H., Allred W. P. and Spitzer, W. G., J. Appl. Phys. **43**, 4096 (1972).
21. Norris B. and Narayanan, G. H., J. Appl. Phys. **48**, 2784 (1977).
22. Pearton, S. J., Kahn, J. M., Hansen, W. L. and Haller, E. E., J. Appl. Phys. **55**, 1464 (1984).
23. Pearton, S. J. and Tavendale, A. J., J. Appl. Phys. **54**, 1375 (1983).
24. Auret, F. D. and Nel, M., Appl. Phys. Lett. **48**, 130 (1985).
25. Guislain, H. J., De Wolf, L. and Clauws, P., J. Electron. Mater. **7**, 83 (1978).
26. Janzén, E., Linnarsson, M., Monemar, B. and Kleverman, M., Mat. Res. Symp. Proc. **163**, 169 (1990).
27. Auret, F. D., Leitch, A. W. R. and Vermaak, J. S., J. Appl. Phys. **59**, 158 (1986).
28. Leitch, A. W. R., Prescha, Th. and Weber, J., Phys. Rev. **B44**, 1375 (1991).
29. Yuan, M. H., Wang, L. P., Jin, S. X., Chen, J. J. and Qin, G. G., Appl. Phys. Lett. **58**, 925 (1994).

Effect of the Interatomic Si-Si-potential on Vacancy Production during Ion Implantation of Si

K. Nordlund, J. Keinonen and A. Kuronen

Accelerator Laboratory, University of Helsinki, P.O. Box 43, FIN-00014, Finland

Received May 3, 1994; accepted June 15, 1994

Abstract

Collision cascades in crystalline silicon due to impinging 10 eV – 1 keV Si atoms are simulated using molecular dynamics methods. The simulations are carried out for 30–100 events to obtain representative statistics for production of different types of vacancies. The results are used to examine the dependence of vacancy production on the interatomic Si-Si potential between the colliding atoms. The dependence of the number of vacancies was found to be sensitive to the form of the potential well but not to the repulsive potential. The results suggest that within the heavily damaged volume of the collision cascade an interatomic potential with somewhat narrower well than that of the commonly used Stillinger-Weber potential should be used to simulate the vacancy production in silicon.

1. Introduction

Processes that produce damage in crystalline silicon during ion implantation have been studied extensively, see e.g. Refs [1–6]. Simulations of vacancy production processes in solids have been carried out using various binary-collision-approximation (BCA) and molecular dynamics (MD) simulation methods [7]. Only rough estimates on the number of vacancies can be obtained with the BCA methods in the framework of the empirical Kinchin-Pease equation [7]. Molecular dynamics simulation methods are required in realistic simulations of collision cascades.

In MD simulations, defect formation has typically been estimated on the basis of a single or a couple of cascade simulations (see e.g. Refs [8, 9]). However, this leads to highly unreliable results, as illustrated by Fig. 1. In Fig. 1 the number of vacancies in ten different collision cascades produced by a 300-eV Si atom recoiling in crystalline silicon (c-Si) are shown as a function of time along with the average value calculated from 100 events (see below).

In previous works where processes in c-Si have been simulated with MD methods employing classical potentials, the number of defects obtained has been found to be significantly higher than experimental values [10]. Low-energy processes in small atom clusters can be more realistically simulated using *ab initio* MD methods [11, 12] than with classical MD methods. However, *ab initio* MD simulations are still far too slow to handle processes involving interaction energies higher than a few eV or systems with more than a few hundreds of atoms. Therefore, a better understanding of classical MD simulation methods is desirable to obtain more realistic defect concentrations in simulations of collision cascades, which involve keV energies and several thousands of atoms.

This work was undertaken to study how the vacancy production during the implantation of 10 eV–1 keV Si atoms into c-Si depends on the interatomic Si-Si potential.

2. Principles of the simulations

The MD simulations were carried out using a modified version of the computer code used previously in our laboratory for the simulation of slowing down of low-velocity recoils (energy less than 100 eV/amu) produced in thermal neutron capture reactions [13].

The potential employed in the simulations was the Stillinger-Weber three-body potential commonly used in simulating the structure of c-Si [14]. To obtain a realistic potential at small separations ($r < 1.7 \text{ \AA}$), the Stillinger-Weber potential was splined between $r = 1.7 \text{ \AA}$ and 2.0 \AA with the repulsive potential.

The simulation of a collision cascade was initiated by selecting a recoiling atom among silicon atoms in one corner of the simulation cell, and giving it a recoil velocity in an isotropically chosen direction in one quadrant. A recoil energy of 1 keV was selected to test the effect of the repulsive potential. To test the effect of the potential well separately, lower energies were chosen. However, in order to obtain statistically significant results it is important that more than only a few vacancies are produced for one recoil event. Therefore, a recoil energy of 300 eV which was found to yield between 20 and 50 vacancies in one recoil event was chosen.

The size of the simulation cell was set to be large enough to contain the entire collision cascade. For 1 keV recoils this amounted to a cell of $76 \times 76 \times 76 \text{ \AA}^3$ with 21952 atoms and for 300 eV recoils to a cell of $54 \times 54 \times 54 \text{ \AA}^3$ with 8000 atoms.

The initial velocities of the atoms in the simulation cell were chosen randomly according to the Maxwell velocity distribution. Thermal movement of the atoms in the cell was

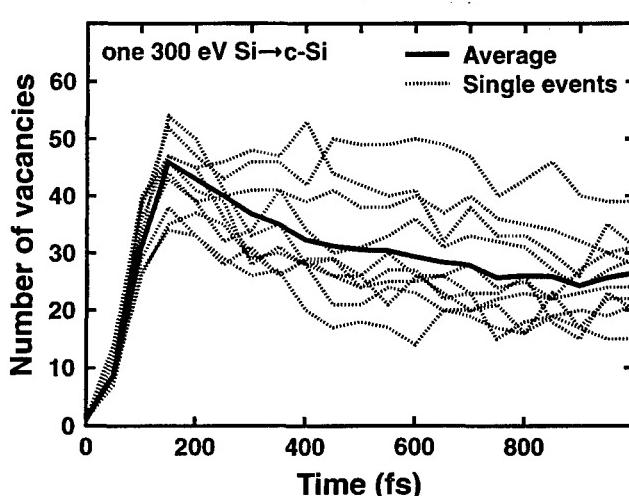


Fig. 1. Number of vacancies produced by 300 eV recoils with randomly selected initial recoil directions. Ten individual events and the average calculated from 100 events are shown.

simulated for 100 fs at 300 K to obtain realistic thermal displacements.

The time step in the simulations was initially 0.1 fs and was made longer during the simulation as the recoil atom slowed down. The longer time steps were selected so that the fastest atom in the simulation cell did not move more than 0.1 Å during one time step. The algorithm employed to solve the equations of motion was a modified Beeman algorithm presented in Ref. [15].

The temperature of the outermost 3 atom layers was scaled down to 300 K at regular intervals to realistically dissipate energy from the simulation cell. To prevent the entire cell from starting to move due to extra momentum obtained from the recoil atom, the sum of the vector momenta of the atoms in the outermost layer was also scaled down to zero. The scalings also prevented kinetic energy from being reflected back from borders of the simulation cell.

Vacancies were detected in several ways during the simulations. The total number of vacancies was calculated by counting all original lattice sites that were empty. For each located empty site the nearest-neighbour sites were examined for emptiness, yielding the numbers of mono-, di-, tri- etc. vacancies. A site was defined as empty if it did not contain any atom within a radius of 1.2 Å, corresponding to the half of the nearest-neighbour distance.

The number of atoms with energies greater than a threshold energy of 15 eV (the minimum value of E_d in the Kinchin-Pease equation, see below) was also calculated.

3. Results and discussion

The dependence of the number of vacancies on the repulsive potential was examined by simulating 50 1-keV recoil vents for three different potentials. The potentials used were the universal ZBL potential [16], a modified Molière potential [17], and a potential obtained from *ab initio* calculations using the commercial DMol program [18, 19]. The potentials are shown in Fig. 2, and the results in Fig. 3.

Although the number of secondary recoils with energies greater than 15 eV (corresponding to the minimum energy required for the formation of a Frenkel pair in c-Si [20, 21]) differ significantly for the three potentials the number of vacancies produced does not show any statistically significant dependence on the repulsive potential.

To test the effect of the form of the potential well on vacancy production, the parameters S , p , m_1 and m_2 in the

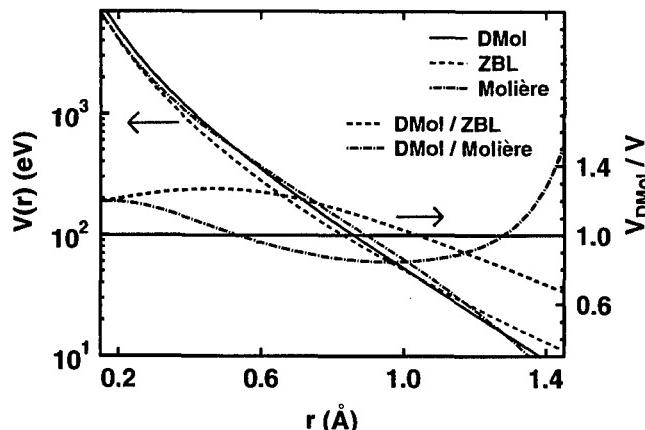


Fig. 2. Interatomic repulsive Si-Si potentials used in the simulations of vacancy production.

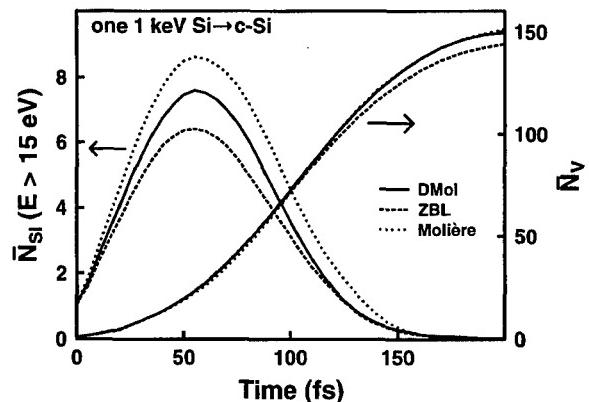


Fig. 3. Average numbers of recoiling Si atoms with energies greater than 15 eV ($\bar{N}_{Si}(E > 15 \text{ eV})$) and number of vacancies (\bar{N}_v) produced in the slowing down process of one 1-keV Si atom in c-Si. The numbers are averages calculated from 50 simulation events for the three potentials shown in Fig. 2.

two-body-part of the Stillinger-Weber-potential

$$V_2(x) = \begin{cases} S\epsilon A(Bx^{-p} - m_1)e^{m_2/(x-a)} & 0 < x < a \\ 0 & x > a, \end{cases} \quad (1)$$

were modified. The unmodified values for the parameters are $S = m_1 = m_2 = 1$ and $p = 4$.

The modified values were selected so that the middle point of the well is at the equilibrium separation of 2.35 Å, and the value of r where the potential crosses x-axis is about the same for all the potentials. The parameter values were selected to yield potentials where either the width, the depth or the product of width and depth was the same as for the unmodified potential. For all the modified potentials the lattice stability was tested by carrying out the simulation of thermal motion at 300 K, and comparing the radial density functions after the simulation. The repulsive potential used in the simulations was the DMol potential.

The simulations were carried out with the unmodified and six modified potentials. Parameters for the width and the depth of the potentials are given in Table I. The width is defined as the full width at half minimum of the total potential, i.e. the potential where the repulsive part has been joined to the modified Stillinger-Weber potential.

The time evolution of the number of vacancies is shown in Fig. 4. The statistical error of the number of vacancies is about 2 for all the potentials. During the first 100 fs the number of vacancies rises about equally for all the potentials. After this the results start to differ, reflecting the differences in the attractive potential. After about 500 fs an approximately stable value in the number of vacancies is reached. Simulations carried out with the unmodified potential for 3000 fs showed that the value reached at 500 fs remained stable.

The number of vacancies, calculated as the average of the values between 500 and 1000 fs, is shown in Table I for all the potentials. Also shown are the number of monovacancies and the sum of the numbers of di-, tri-, tetra- and pentavacancies, calculated as the average of the values between 800 and 1000 fs.

For the unmodified Stillinger-Weber potential (potential 1) the number of vacancies (N_v) is 27. Potential 2 with the same depth and a 5% narrower width, results in a significantly smaller number of vacancies than potential 1, i.e. 16. Potential 3 which has the same width as potential 2 but where the depth has been modified so that the product of

Table I. Widths and depths for the unmodified and six modified Stillinger-Weber potentials. The average number of vacant sites, monovacancies and more complex vacancies produced in a collision cascade by one 300 eV Si atom recoiling in c-Si are shown in the three last columns

Potential	Width (Å)	Depth (eV)	Vacancies	Monovac.	Others
1. (Unmodified)	0.903	2.16	27	15	8
2. (narrower)	0.847	2.16	16	12	4
3. (narrower, deeper)	0.847	2.28	19	13	6
4. (wider)	0.924	2.16	37	18	15
5. (wider, shallower)	0.924	2.10	36	19	12
6. (deeper)	0.903	2.28	32	15	12
7. (shallower)	0.903	2.10	23	13	7

the width and depth is the same as for potential 1, yields about the same value of N_V as potential 2.

Potential 4 which has the same depth as potential 1 but a 2% wider width, leads to a clearly higher value of $N_V = 37$ than potential 1. Potential 5 with the same width as potential 4 and the same product of width and depth as potential 1, results in about the same number of vacancies as potential 4.

In potentials 6 and 7 the depth of the potential has been modified to the same values as for potentials 3 and 5, respectively, but the width is the same as for potential 1. The number of vacancies obtained differs somewhat from the values of the unmodified potential but the difference is significantly smaller than for the potentials 2, 3, 4 and 5 where the width was changed.

The results indicate that the width of the potential has a strong effect on N_V . The behaviour can be understood qualitatively in terms of the interatomic forces. The width of the well reflects the strength of the forces involved when the bond is about to be broken. A narrower well means that the derivative of the potential, i.e. the force between the atoms, is stronger at the well borders resulting in stronger bonds and thus smaller numbers of vacancies.

Comparison of the potentials 3 and 5 with the potentials 2 and 4, respectively, and the potentials 6 and 7 with the potential 1 show that a deeper potential yields a somewhat higher number of vacancies, and vice versa. This somewhat surprising effect is significantly weaker than the effect of the width of the potential, however.

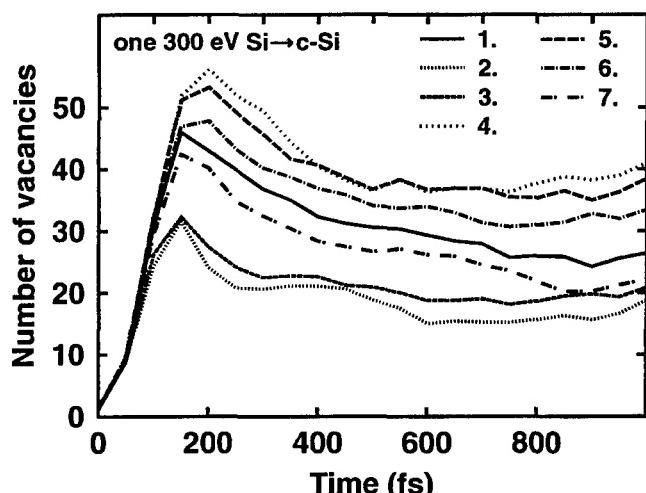


Fig. 4. Average number of vacancies produced during the first 1000 fs by one 300 eV Si atom for the 7 modified Stillinger-Weber potentials given in Table I. The numbers are averages calculated from at least 30 simulation events.

What is especially interesting is that relatively small changes in the potential result in large changes in N_V . For instance, a 2% increase in the width of the potential between the potentials 1 and 4 yields a 37% increase in N_V .

The time evolution of different vacancy type defects is shown over a longer time scale in Fig. 5. The number of pentavacancies is negligible and is therefore not shown. The number of complex (di-, tri-, and tetra-) vacancies increases rapidly in the beginning of the cascade, but decreases slowly after that, whereas the number of monovacancies increases steadily. This indicates that the complex vacancies slowly break down to monovacancies.

Because monovacancies are mobile in silicon at room temperature [22], they should disappear on a longer time scale, of the order of μs or ms . Due to limitations in the available computer capacity, simulations of this process could not be carried out.

The Kinchin-Pease equation is generally used to estimate the number of vacancies produced in ion irradiations of solids. The total number of vacancies N_V produced during ion irradiation is

$$N_V = \frac{F_{D_n}}{2E_d} \quad (2)$$

where F_{D_n} is the deposited nuclear energy.

Values for the parameter E_d have been determined empirically by measuring the number of Frenkel pairs produced. The values of E_d given in the literature for silicon range from about 15 to 25 eV, Refs [20, 21]. The equation predicts that the number of Frenkel pairs (corresponding to the number of monovacancies in our simulations) produced

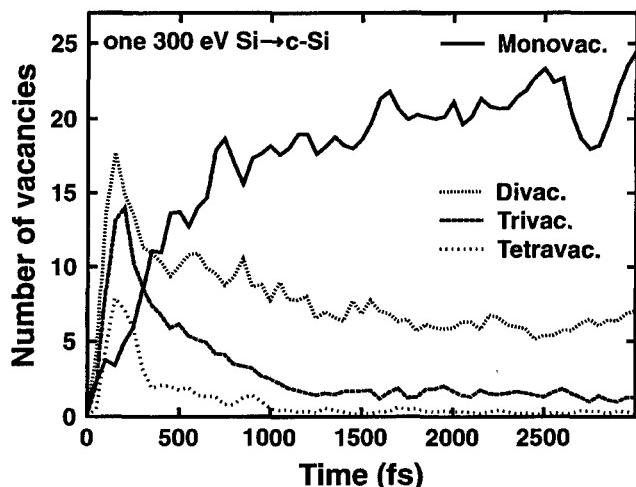


Fig. 5. Time evolution of the number of mono-, di-, tri- and tetravacancies. The numbers are averages calculated over 10 events with the unmodified Stillinger-Weber potential.

by one 300-eV recoil should be below 10. Comparison of this value with the values for monovacancies in Table I indicate that the unmodified Stillinger-Weber potential is not suitable for simulations of collision cascades. The numbers of vacancies obtained with the potentials with a narrower well are nearer the values predicted by the Kinchin-Pease equation.

4. Conclusions

In this work we demonstrated the need and feasibility to collect representative statistics when vacancy production in collision cascades due to ion irradiations is simulated in MD calculations. We showed that vacancy production is sensitive to the attractive potential but not to the repulsive potential. We obtained that the Stillinger-Weber potential suitable for the simulations of c-Si at or near equilibrium does not adequately describe interatomic interactions in collision cascades and that a modified potential with an at least 5% narrower well than that of the Stillinger-Weber potential results in a realistic description of vacancy production in classical MD simulations.

References

1. Lindner, J. K. N., Nucl. Instr. Meth. Phys. Res. **B84**, 153 (1994).
2. Campisano, S. U., Coffa, S., Rainieri, V., Priolo, F. and Rimini, E., Nucl. Instr. Meth. Phys. Res. **B80/81**, 514 (1993).
3. Tsai, C. C., Anderson, G. B. and Thompson, R., Mat. Res. Soc. Symp. Proc. **192**, 475 (1990).
4. Tsao, J. Y., Chason, E., Horn, K. M., Brice, D. K. and Picraux, S. T., Nucl. Instr. Meth. Phys. Res. **B39**, 72 (1989).
5. Hasan, M.-A. et al., J. Appl. Phys. **65**, 172 (1989).
6. Erola, M., Keinonen, J., Hautala, M. and Uhrmacher, M., Nucl. Instr. Meth. Phys. Res. **B34**, 42 (1988).
7. Beeler, J. R., "Radiation effects computer experiments" (North Holland, Amsterdam 1983).
8. Zhu, H. and Averback, R. S., Nucl. Instr. Meth. Phys. Res. **B83**, 334 (1993).
9. Diaz de la Rubia, T. et al., Nucl. Instr. Meth. Phys. Res. **B80/81**, 86 (1993).
10. Drabold, D. A., Fedders, P. A., Sankey, O. F. and Dow, J. D., Phys. Rev. **B42**, 5135 (1990).
11. Car, R. and Parrinello, M., Phys. Rev. Lett. **60**, 204 (1988).
12. Fedders, P. A., Drabold, D. A. and Klemm, S., Phys. Rev. **B45**, 4048 (1992).
13. Keinonen, J. et al., Phys. Rev. Lett. **67**, 3692 (1991).
14. Stillinger, F. H. and Weber, T. A., Phys. Rev. **B31**, 5262 (1985).
15. Smith, R. and Harrison, D. E., Computers in Physics Sep/Oct 1989, 68 (1989).
16. Ziegler, J. F., Biersack, J. P. and Littmark, U., in "The Stopping and Range of Ions in Matter" (Pergamon, New York 1985), vol. 1.
17. Eckstein, W., in Ref. [23], eq. (4.4.3) on p. 55.
18. Keinonen, J., Kuronen, A., Nordlund, K., Nieminen, R. M. and Seitsonen, A. P., accepted for publication in Nucl. Instr. Meth. Phys. Res. B.
19. DMol is a trademark of Bio Sym. Inc., San Diego, California, USA.
20. Narayan, J., Fath, D., Oen, O. S. and Holland, O. W., J. Vac. Sci. Technol. **A2**, 1303 (1984).
21. Lannoo, M. and Bourgoin, J., "Point Defects in Semiconductors" (Springer, Berlin 1981), vol. II, p. 131.
22. Lannoo, M. and Bourgoin, J., "Point Defects in Semiconductors" (Springer, 1981), vol. II.
23. Eckstein, W., "Computer Simulations of Ion-Solid Interactions" (Springer-Verlag, Berlin 1991), p. 40.

Surface Morphology Study of Titanium Silicide Formed on Polycrystalline Silicon

Tord E. Karlin¹, Wlodek Kaplan^{1,2} and Shili Zhang¹

¹ Royal Institute of Technology, Solid State Electronics, P.O. Box E229, S-164 40 Kista-Stockholm, Sweden

² Industrial Microelectronics Center, P.O. Box 1084, S-164 21 Kista-Stockholm, Sweden

Received June 13, 1994; accepted in revised form July 12, 1994

Abstract

TiSi_2 can be used as a shunting layer to decrease the resistance of the poly-Si lines used as interconnections in integrated Si technology. In this work the correlation between the silicide thickness, the silicide surface roughness, the silicide grain size, the dopants in Si, and the crystallinity of Si was investigated on unpatterned wafers. The results will be used as reference in a further study of silicide formation on sub-micron poly-Si lines. After silicide formation at 700 °C, the surface roughness has been found to be strongly dependent on the thickness of the silicide layer formed. The dopants in the Si did not only retard the formation of the C49 phase, but also hindered the transformation of TiSi_2 from the C49 to the low resistive C54 phase. The surface roughness did not change when a second annealing step at 850 °C was used to transform the silicide from the C49 phase to the C54 phase.

1. Introduction

Metallisation issues become critical when the width of the MOSFET channel length as well as the line width of the interconnections are scaled down to sub-micron dimensions. Narrow polycrystalline silicon (poly-Si) lines suffer from high resistance which leads to high voltage drops and long RC delay times. It is therefore crucial to find another material with lower resistivity to replace poly-Si. With the use of multi-level metallisation schemes the material also has to stand high process temperatures. A thin refractory silicide film on top of the poly-Si (so called polycide) can fulfil these requirements. Titanium disilicide has been the most commonly used silicide for this purpose [1].

TiSi_2 has two different phases [2, 3]. At low formation temperatures (400 °C to 700 °C) a phase (C49) is formed with a resistivity around 65 $\mu\Omega\text{cm}$. At higher temperatures the C49 phase is transformed to the C54 phase with a resistivity around 15 $\mu\Omega\text{cm}$, only five times higher than the resistivity of Al. It is the C54 phase that is required for gate and interconnection applications. The morphology of thin and narrow lines and silicide characteristics such as grain size, surface roughness, nucleation and thermal stability have to be carefully controlled.

It has been reported that the transformation of TiSi_2 from the C49 to the C54 phase is affected by the dopants in Si [4, 5], the silicide thickness [6] and the line width [7, 8]. In this study the correlation between the surface roughness, the silicide grain size, the silicide thickness, the dopants in Si, and the crystallinity of Si is investigated on unpatterned wafers. The results will be used as reference in a further study of silicide formation on sub-micron poly-Si lines.

2. Experimental

The Ti silicide was formed by reaction with 4000 Å thick Si layers, deposited on top of a 1000 Å SiO_2 . The silicon layers

were deposited by LPCVD at 560 °C and 625 °C, resulting in amorphous (α -) Si and poly-Si, respectively. The deposited layers were either doped by ion implantation with As or B, or undoped. The implantation dose was $1 \cdot 10^{16} \text{ cm}^{-2}$ for both As and B. All wafers were heat treated at 900 °C for 60 minutes in O_2 to activate the dopants. The undoped wafers were subjected to the same heat treatment. During the dopant activation, the α -Si layers became crystallised, but for the sake of convenience " α -Si" will still be used throughout this work, to identify the crystallised α -Si layers. A single layer of Ti, 630 Å thick, or a bilayer consisting of Ti, 400 Å thick with a TiN cap, 800 Å thick, was deposited on the Si by sputtering. On wafers without the TiN cap the Ti was especially thick in order to compensate for the expected nitridation during silicide formation, which was performed in N_2 . The TiN cap layer was used to prevent severe interaction between Ti and N_2 , so as to reduce the surface roughness. A two-step silicide formation process [9] was used to form the silicide. After the first step, by rapid thermal annealing (RTA) in N_2 at 700 °C for 30 s, TiSi_2 of the C49 phase was expected to form. The TiN (deposited or formed during silicidation) and the unreacted Ti were removed from all wafers with a standard etch sequence consisting of etching in a solution of $\text{H}_2\text{SO}_4 : \text{H}_2\text{O}_2 = 4 : 1$ followed by etching in another solution of $\text{NH}_4\text{OH} : \text{H}_2\text{O}_2 : \text{H}_2\text{O} = 1 : 1 : 5$ [9]. The wafers were then subjected to a second RTA at 850 °C for 30 s in N_2 to transform the TiSi_2 from the C49 phase to the low resistive C54 phase. The experimental conditions are summarized in Table I.

X-ray diffractometry (XRD) was used to identify the silicide phase and to determine the approximate grain size. Rutherford backscattering spectrometry (RBS) with ${}^4\text{He}^+$ ions at 2.4 MeV was employed to measure the thickness of the silicide and the remaining silicon layers. The resistivity of the two different silicide phases was calculated from the thickness determined by RBS and the sheet resistance measured by a 4-point probe. The surface roughness of the silicides was measured with an atomic force microscope (AFM). The root mean square (RMS) roughness is used as a figure of merit for the surface roughness [10].

3. Results and discussion

3.1. Silicide formation

Figure 1 shows that the dopant in the Si retards the silicide formation during the first RTA step at 700 °C. The n-type dopant (As) affects the formation more than the p-type dopant (B). The silicide formed on the uncapped wafers is thicker than that of the wafers with the TiN cap, probably

Table I. Experimental conditions

Si crystallinity as-deposited	Si doping	Ti thickness as-deposited [Å]	TiN thickness as-deposited [Å]	TiSi ₂ thickness as-formed [Å]
α-Si	—	630	—	1030
α-Si	—	400	800	580
α-Si	Boron	630	—	830
α-Si	Boron	400	800	680
α-Si	Arsenic	630	—	780
α-Si	Arsenic	400	800	630
poly	—	630	—	1130
poly	—	400	800	780
poly	Boron	630	—	830
poly	Boron	400	800	680
poly	Arsenic	630	—	760
poly	Arsenic	400	800	580

because of less nitridation of the uncapped titanium than expected.

To further investigate the influence of doping and Si crystallinity on the silicide formation, a test was performed where the as-deposited amorphous silicon was not subjected to the activation anneal even after dopant implantation, thereby avoiding Si crystallisation. Silicide was formed on As doped and undoped wafers by reacting a 400 Å thick Ti layer, capped with 500 Å of TiN, at 650 °C for 60 s in N₂. On the undoped amorphous silicon a 870 Å thick silicide was formed. The silicide layer was partially transformed to the C54 phase already at 650 °C. On the As doped amorphous Si, however, the silicide was only 340 Å thick but the C49 phase alone could be detected. An amorphous Si layer has a high enthalpy which could enhance silicide formation [11]. The presence of As effectively slows the formation because the As atoms segregate the grain boundaries of the C49 TiSi₂, thereby reducing Si transport along these high-diffusion paths [4, 5].

The surface roughness and the silicide grain size are strongly dependent on the silicide thickness, as depicted in Figs 2 and 3, respectively. To describe different wafers the following convention is used. “α-Si” and “poly” represent the Si substrate layers. “—”, “B” and “As” indicate the Si doping: undoped, B doped and As doped respectively. “Ti” and “Ti/TiN” represent the uncapped and capped wafers. The surface roughness is also affected by the type of Si

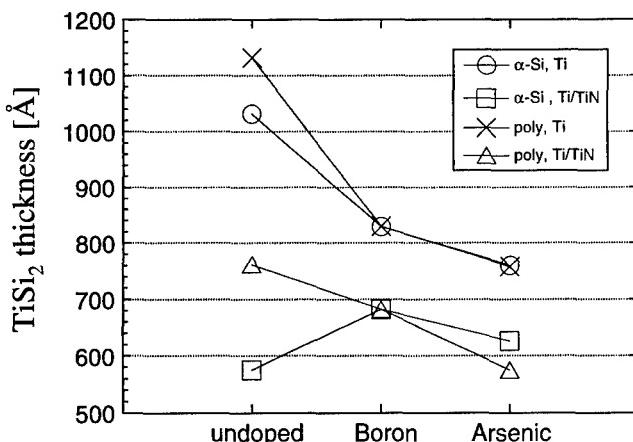


Fig. 1. Impact of dopants (As and B) on the thickness (measured with RBS) of the TiSi₂ formed on α-Si and poly-Si substrates, with and without the TiN cap layer. The experimental conditions for preparation of the wafers are given in Table I.

doping (e.g. wafer {α-Si, B, Ti} vs. wafer {α-Si, As, Ti} and wafer {α-Si, B, Ti/TiN} vs. wafer {α-Si, As, Ti/TiN} in Fig. 2). In general, As doped Si gives a rougher silicide surface than B doped Si. The difference in silicide surface roughness between the α-Si and the poly-Si wafers was negligibly small (e.g. {α-Si, B, Ti/TiN} vs. {poly, B, Ti/TiN} and {α-Si, As, Ti} vs {poly, As, Ti}), although it was observed that after dopant activation and before Ti deposition, the As doped α-Si had a smoother surface than the As doped poly-Si. The RMS surface roughness for the As doped α-Si was 130 Å while it was 230 Å for the As doped poly-Si. Furthermore,

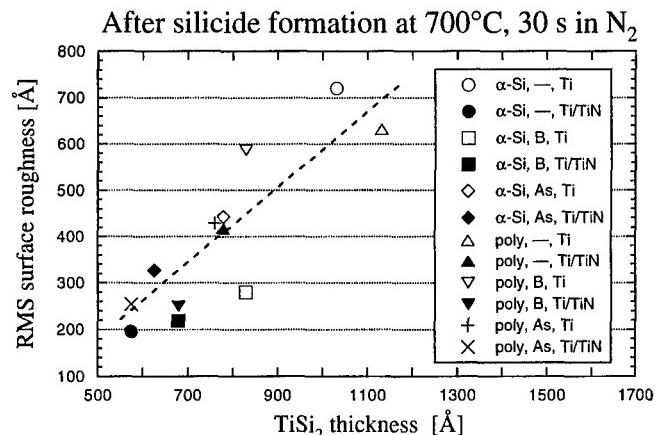


Fig. 2. Dependence of the RMS roughness values (determined by AFM) on the thickness of the silicide formed at 700 °C.

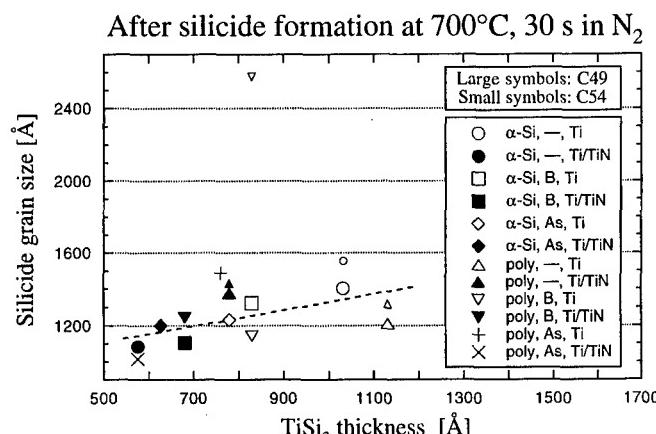


Fig. 3. Dependence of the silicide grain size (estimated by XRD) on the thickness of the silicide formed at 700 °C.

after the activation the preferential orientation of the silicon layers was (200) for all wafers with poly-Si and (111) for all wafers with crystallized α -Si. The big difference in surface roughness between wafer {a-Si, B, Ti} and wafer {poly, B, Ti} is attributed to the formation of C54 phase, which will be discussed later.

Since the silicide formed on the wafers capped with the TiN was thinner than that on the wafers without the cap, it is difficult to observe the influence of the TiN cap on the silicide surface roughness. The impact of the TiN cap is probably a minor effect compared to the predominant influence of the silicide thickness.

3.2. C49 to C54 phase transformation at 700 °C

The XRD spectra in Fig. 4 show that the transformation from the C49 to C54 phase has taken place at 700 °C for the silicide formed on the undoped wafers except that with the α -Si and the TiN cap. On the doped wafers, a small amount of C54 phase was detected only on the B doped poly-Si wafer without the TiN cap. No C54 phase was detected for the equally thick silicide formed on B doped α -Si without the TiN cap. It is observed in Fig. 3 that the grain size of C49 and C54 is comparable for the undoped wafers { α -Si, —, Ti}, {poly, —, Ti} and {poly, —, Ti/TiN}. However, for wafer {poly, B, Ti}, the grain size of the small amount of C54 phase is much larger even though the silicide is much thinner than that on wafers {a-Si, —, Ti} and {poly, —, Ti}.

The presence of dopants seems to have not only retarded the formation of the C49 phase but also hindered the nucleation of the C54 phase. The latter implies a much lower density of nucleation sites for the C54. However, once a nucleus of the C54 phase forms and begins to grow, it will grow into a big grain before it encounters any competing grains. This leads to a rougher surface (wafer {poly, B, Ti}) as compared to the case without the transformation (wafer { α -Si, B, Ti}), as observed in Fig. 2.

3.3. Surface morphology

AFM pictures of the silicide surfaces after formation at 700 °C showed a distinct difference in surface morphology between wafers with a small amount of C54 and wafers where only the C49 phase was detected. Figure 5(a) shows the surface, measured by AFM, for wafer { α -Si, B, Ti} with about 800 Å of silicide after the first RTA. The silicide consisting of pure C49 phase according to the XRD spectra, has

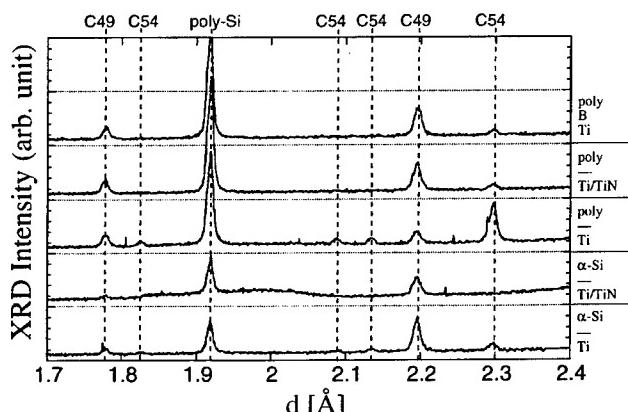


Fig. 4. X-ray diffraction patterns for the undoped wafers and for a doped wafer on which a small amount of C54 phase is observed (wafer {poly, B, Ti}). No C54 phase was detected for the other doped wafers.

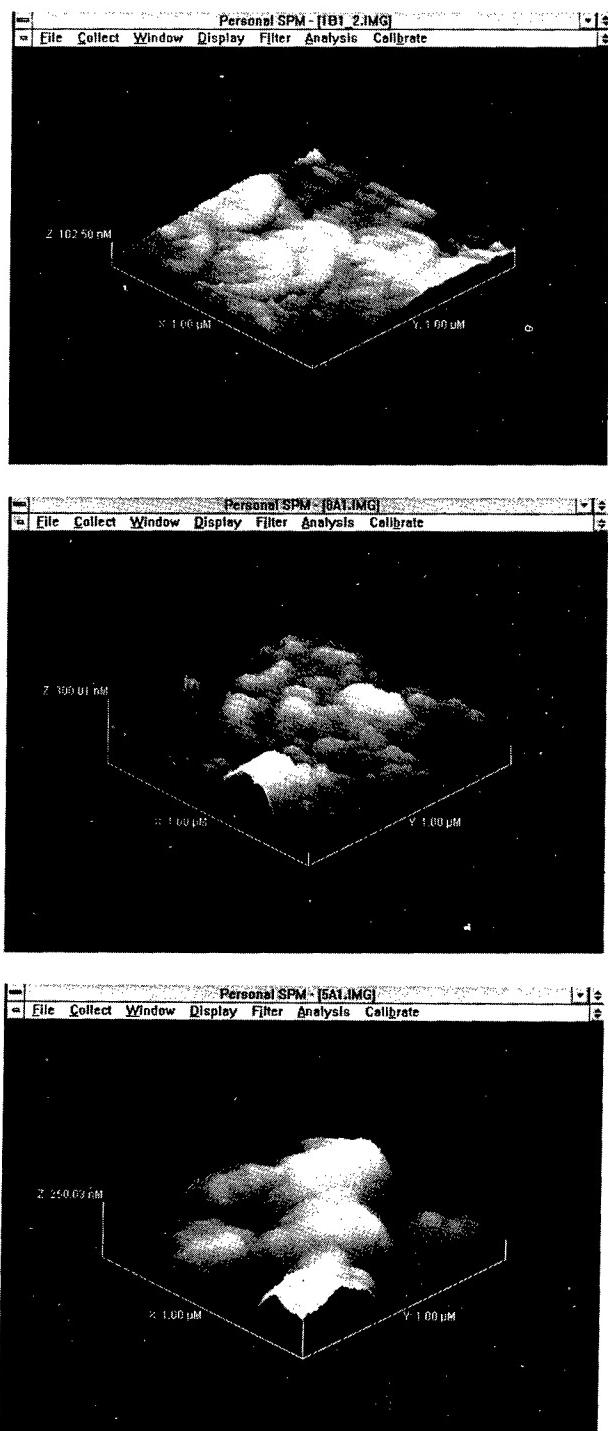


Fig. 5. AFM pictures showing (a) a smooth surface of wafer { α -Si, B, Ti}; (b) a rougher surface of wafer {poly, B, Ti} consisting of a few large hills; and (c) a very rough surface of wafer { α -Si, —, Ti} mainly consisting of large hills.

a relatively smooth surface. Figure 5(b) shows the surface of the equally thick silicide on wafer {poly, B, Ti}. Now the silicide has been partially transformed to C54 according to the XRD results (Fig. 4). This surface is considerably rougher with some large hills surrounded by smaller hills. The large hills are most probably composed of the C54 phase. Figure 5(c) shows the surface of wafer { α -Si, —, Ti}. More C54 phase has been formed for this wafer, according to XRD (Fig. 4). The surface is very rough with mainly large hills, consisting presumably of the C54 phase.

The resistivity of the C49 silicide phase was 65–95 $\mu\Omega\text{cm}$ after the first anneal at 700 °C. After the second anneal at

850 °C, the resistivity of the C54 phase was only 15–25 $\mu\Omega\text{cm}$, indicating that the silicide layers were still continuous after transformation to the C54 phase. AFM data showed that the surface roughness was almost intact after the second anneal at 850 °C, which is in agreement with our earlier work where 900 °C was used for the second anneal [12].

For device applications, a rough surface with large crystallite grains is not desired, particularly for deep sub-micrometer metallisations, where the line width ($\approx 0.2 \mu\text{m}$) is comparable with the grain size ($\approx 0.1\text{--}0.3 \mu\text{m}$). Therefore, the annealing temperature used for the first step RTA should be kept below a certain threshold in order to avoid the formation of the C54 phase. The upper temperature limit could be set at 700 °C, according to the present study. The first annealing temperature is also limited by the silicide formation on SiO_2 which leads to bridging between the gate and source/drain areas of a MOSFET [9]. However, in order for the transformation to take place easily, a relatively thick TiSi_2 layer of the C49 phase is desired [6]. This sets the lower temperature limit for the formation. For example, forming the silicide at 650 °C for 60 s resulted in approximately 30% less silicide than when using 700 °C for 30 s [13].

The transformation anneal during the second step RTA should be performed at a relatively high temperature in order to create a high density of nucleation sites, thereby preventing the silicide grains from growing large. Our earlier work showed that a transformation anneal at 900 °C did not increase the roughness of the silicide surface [12]. However, the temperature for the second step RTA is limited by the thermal stability of TiSi_2 on poly-Si. Agglomeration of the silicide layer, which is also thickness dependent [14], should be avoided. The upper temperature limit for agglomeration of TiSi_2 on poly-Si has experimentally been shown to be around 850 °C–900 °C [15, 16]. The thermal stability of TiSi_2 on poly-Si is somewhat enhanced when the poly-Si is doped [12, 17]. Thus, the temperature windows for the formation and transformation are very narrow. Lasky *et al.* have shown [7] that the temperature limit for agglomeration of TiSi_2 on narrow lines is 950 °C–1000 °C. On the other hand the formation of TiSi_2 on sub-micron wide poly-Si lines would also demand a higher temperature [7,8] than what have been used in this study, both for the formation and for the transformation, thereby shrinking the temperature window even further.

4. Conclusions

The surface morphology of the TiSi_2 formed on p-type, n-type and undoped poly-Si and crystallized α -Si has been

studied. After silicide formation at 700 °C, the surface roughness has been found to be strongly dependent on the thickness and the phase of the silicide formed. The influence of the silicon crystallinity on the silicide surface roughness was negligible. The dopants in the Si not only retard the formation of the C49 phase, but also hinder the transformation of the C49 to the C54 phase. In order to avoid the nucleation of the C54 phase during the first step RTA, the formation temperature should be kept low (probably below 700 °C for B doped silicon). However, a low formation temperature results in thin silicide layers, which in turn retards the transformation during the second step RTA. The required transformation temperature might then be well above 900 °C, which can lead to silicide agglomeration. The temperature windows for the formation and transformation have become very narrow, and tend to become infinitely so when forming the silicide on sub-micron poly-Si lines.

Acknowledgements

The authors are grateful to Prof. F. M. d'Heurle for enlightening discussions. They would like to thank N. Lundberg for the RBS analysis, and to acknowledge B. Rodell, J. Wittborn and Prof. K. V. Rao for invaluable help with the AFM analysis. This work was supported by the Swedish Board for Technical Development (NUTEK).

References

1. Ting, C. Y., d'Heurle, F. M., Iyer, S. S. and Fryer, P. M., *J. of Electrochem. Soc.* **133**, 2621 (1986).
2. d'Heurle, F. M. *et al.* "The two Crystalline Structures of TiSi_2 : Identification and Resistivity" (IBM, T. J. Watson Research Center, Yorktown Heights, 1985), RC 11151.
3. Beyers, R. and Sinclair, R., *J. Appl. Phys.* **57**, 5240 (1985).
4. Révész, P., Gyimesi, J. and Zsoldos, E., *J. Appl. Phys.* **54**, 1860 (1983).
5. Beyers, R., Coulman, D. and Merchant, P., *J. Appl. Phys.* **61**, 5110 (1987).
6. Jeon, H., Sukow, C. A., Honeycutt, J. W., Rozgonyi, G. A. and Nemachich, R. J., *J. Appl. Phys.* **71**, 4269 (1992).
7. Lasky, J. B., Nakos, J. S., Cain, O. J. and Geiss, P. J., *IEEE Trans. Electron Devices* **ED-38**, 262 (1991).
8. Ohguro, T., Morimoto, T., Nishiyama, A., Ushiku, Y. and Iwai, H. *ESSDERC'93*, 481 (1993).
9. Kaplan, W., Zhang, S.-L., Norström, H., Ostling, M. and Lindberg, A., *Microelectronic Engineering* **19**, 661 (1992).
10. Bennett, J. M. and Mattsson, L., "Introduction to Surface Roughness and Scattering" (Optical Society of America, Washington D. C. 1989).
11. d'Heurle, F. M., *J. Mater. Res.* **3**, 167 (1988).
12. Karlin, T. E. *et al.*, *Appl. Surf. Sci.* **73**, 280 (1993).
13. Kaplan, W., Karlin, T. E. and Zhang, S.-L., unpublished results.
14. Nolan, T. P., Sinclair, R. and Beyers, R., *J. Appl. Phys.* **71**, 720 (1992).
15. Ting, T. Y., d'Heurle, F. M., Iyer, S. S. and Fryer, P. M., *J. Electrochem. Soc.* **133**, 264 (1986).
16. Zheng, L. R. *et al.*, *Appl. Phys. Lett.* **48**, 767 (1986).
17. Shenai, K., *J. Mater. Res.* **6**, 1502 (1991).

A Fine Pattern GTO Thyristor Fabricated Using a Self-Aligned Process

M. Bakowski, U. Gustafsson, H. Elderstig, W. Kaplan, P. Norlin and U. Wennström

Industrial Microelectronics Center, P.O. Box 1084, S-164 21 Kista-Stockholm, Sweden

Received May 26, 1994; accepted June 15, 1994

Abstract

GTO thyristors with a 2 cm^2 active area and small rectangular $40 * 40\mu\text{m}$ emitters were fabricated using a specially developed self-aligned process based entirely on reactive ion etching, RIE, technology. Each device consists of 176 1 mm^2 segments containing 100 individual emitters. The devices were electrically evaluated and compared to standard, wet etched mesa, GTO devices of comparable size having $3000 * 200\mu\text{m}$ large emitters. Simulations were carried out to analyse specific features of the new devices. In particular the influence of the surface quality on the gate trigger current and on-state voltage was investigated. It was demonstrated that it is possible to reduce significantly the GTO unit cell dimensions without sacrificing the yield. 23% of the cm^2 devices out of a lot containing 50 devices could be accepted without repair while the overall segment yield was 97%. The dynamic performance of the devices under both snubbered and snubberless conditions compares well to that of the reference devices.

1. Introduction

There is a trend towards smaller emitter dimensions in GTO thyristor design [1–3]. This decrease in unit cell dimensions opens new possibilities to improve the usual trade off between the dynamic and static properties, namely that between maximum turn-off current and gate trigger current. The GTO thyristors are normally designed with a recessed gate which allows devices to be placed in pressure contact type encapsulation. The recessed gate is normally processed by wet etching, producing mesas of emitters surrounded by the gate. The size of the active area of a typical 3000 A device is $20\text{--}30\text{ cm}^2$. The processing of such devices with significantly reduced dimensions would lead to serious problems in photolithographic patterning. In order to be able to fabricate fine pattern GTO devices with the area and yield comparable to those of standard GTO devices, it is crucial to use a self-aligned process. We have been evaluating the concept of fine pattern GTO realised in alternative ways. We call our concept the Compact GTO (CGTO) since we put several individual emitters together in a segment under common cathode metal. The first realisation of the CGTO was based on combined wet etched mesa and planar technologies [4]. This paper reports on a CGTO utilising a self-aligned process and RIE technology. Results from first working devices are reported here. Most importantly, the critical processing steps associated with the technology are identified and their impact on the electrical characteristics is analysed using 2D simulations.

2. Fabrication of devices

Several batches of 2 cm^2 test devices were processed using $390\mu\text{m}$ $120\Omega\text{cm}$ and $550\mu\text{m}$ $200\Omega\text{cm}$ $\langle 111 \rangle$ silicon. Boron implantation was used for the p-base and Accuspin spin-on

doping is used to form the p^+ anode emitter. The n^+ cathode emitter was produced by a POC_3 process. The anode shorts were created by simply masking the Accuspin doping ($390\mu\text{m}$ devices) or by additional POC_3 doping of shorts ($550\mu\text{m}$ devices). The anode shorts pattern was rectangular with an even distribution over the area, and the shorted area was 50% of the active device area. The devices were designed for a blocking voltage of the order of 2kV using a 3.5° negative bevel angle for junction termination. Most of the runs were used for the development and verification of the self-aligned process. Three of the runs resulted in working devices which have been given full electrical evaluation. The layout of the test devices is shown in Fig. 1.

2.1. The self-aligned process

A self-aligned process utilising standard steps from VLSI technology has been developed [5]. The self-aligned processing replaces the photolithography in the most critical patterning steps. These are gate contact definition, emitter

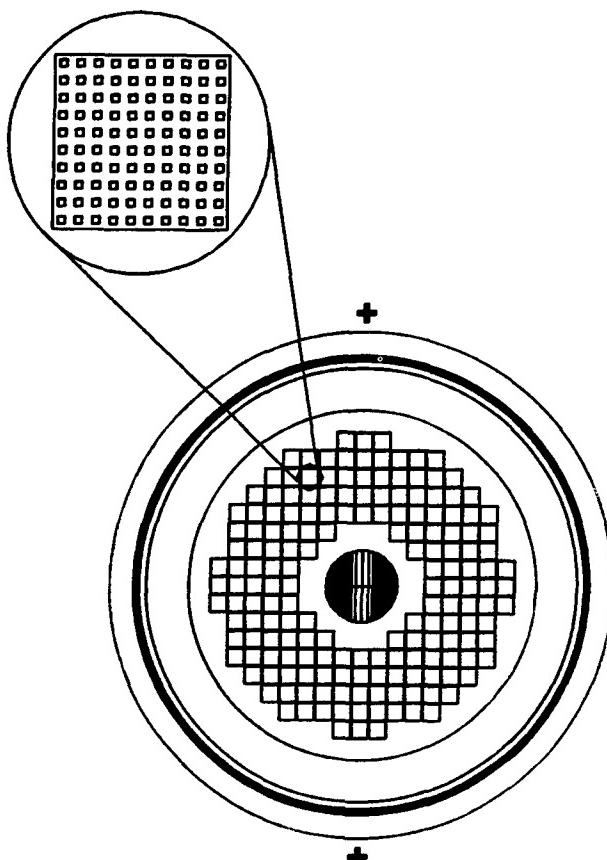


Fig. 1. The layout of the CGTO devices.

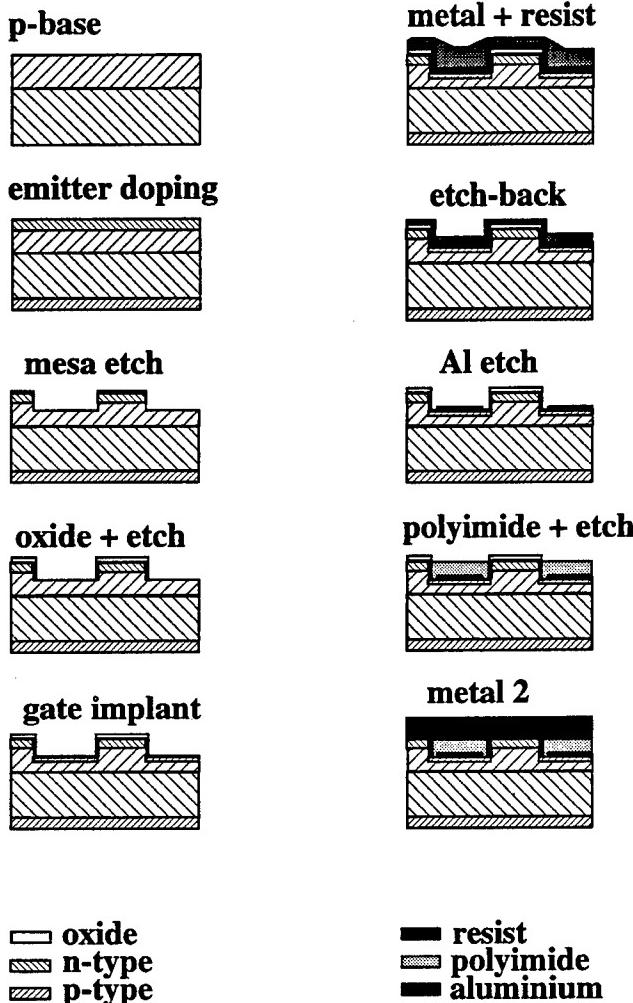


Fig. 2. The main steps of the process sequence.

junction passivation and gate metallisation and isolation. The process sequence is summarised in Fig. 2. Following the p-base implantation and drive-in, Accuspin and POC1₃ doping of the anode and cathode emitters, respectively, are carried out. In the case of the 550 µm thick devices, the anode shorted areas are doped in the same POC1₃ process in which cathode emitters are formed. A 10 µm deep RIE mesa etch of the silicon is performed next to create the cathode emitters. This is followed by a thermal oxidation.

Table I

Device	I_{GT} (mA)	V_T (V)	V_D (kV)
	$V_D = 24 \text{ V}$, $R_L = 0.2 \Omega$ $T_j = 30^\circ\text{C}$	$j_{TQ} = 110 \text{ A/cm}^2$ $T_j = 125^\circ\text{C}$	10 mA $T_j = 125^\circ\text{C}$
CGTO (390 µm)	400	1.95–2.35	1.7
CGTO (550 µm)	260	2.85	2.0
Ref. GTO (390 µm)	250	2.15	2.0

Table II

Device	p-Base resistivity (Ω/sq)	Max. turn-off current density, j_{TQmax} , (A/cm ²)
CGTO (390 µm)	100	200
CGTO (550 µm)	190	—
Ref. GTO (390 µm)	100	190

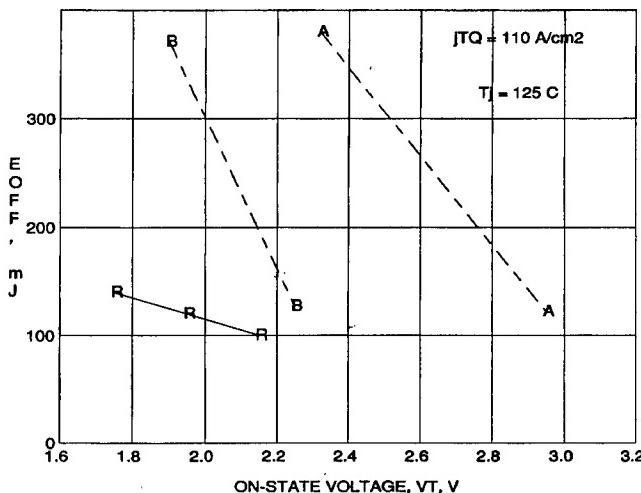


Fig. 3. Turn-off losses per puls versus on-state voltage for the CGTO and references devices. Silicon thickness is 390 µm. The reference devices are electron irradiated. The CGTO devices are as-processed and after proton irradiation.

The first self-aligned processed step uses the fact that thermal oxide is thicker on the n-doped mesas compared to the recessed and lower doped gate area. A subsequent RIE etch back uncovers the gate area. An extra p⁺ doping in the gate is now done in a self-aligned fashion. The oxide spacers on the sides of the mesas and the oxide on top of them are used as a mask during the p⁺ Boron implantation and as the passivation of the cathode junction. The self-aligned patterning of the gate metal uses the fact that photoresist is thicker in the recessed area compared to the mesas. A partial etch-back of the photoresist uncovers the emitters. After subsequent aluminium etch only the metal over the gate area remains. This technique is then repeated with polyimide instead of the photoresist to form the isolation on top of the gate metal. Finally, the cathode metal is formed and patterned in order to create the segmented layout facilitating the repair. The final steps are laser cutting, junction termination and edge passivation.

3. Electrical evaluation

The devices were subjected to thorough electrical evaluation both statically and dynamically. Typical values of the main electrical parameters are shown in Table I and compared to those of the reference devices at a similar value of V_T . The CGTO devices are as-processed while the reference devices are electron irradiated with a dose of about $1.7 \cdot 10^{13} \text{ cm}^{-2}$. The reference devices are factory made standard devices, with the same silicon thickness and resistivity as the CGTO devices. The measurement conditions were as follows: forward blocking voltage was measured at 10 mA leakage current, the on-state voltage was measured at 125 °C, the gate trigger current was measured at $V_D = 24 \text{ V}$ and $R_L = 0.2 \Omega$ in the temperature range 25 to 125 °C. The turn-off losses were measured at 125 °C and with $V_{DM} = V_D = 1200 \text{ V}$, $I_{TQ} = 200, 300$ and 400 A , $C_s = 1 \mu\text{F}$ and $dI_{RG}/dt = -14 \text{ A}/\mu\text{s}$. The snubberless SOA was determined by measuring the maximum turn-off current, I_{TQM} , with $V_{DM} = V_D = 1600 \text{ V}$, $C_s = 0.375 \mu\text{F}$ and $dI_{RG}/dt = -14 \text{ A}/\mu\text{s}$ at 125 °C. Snubberless SOA was determined by measuring the maximum anode sustain voltage, V_{crit} , under inductive load

clamped anode voltage conditions with $j_{TQ} = 100 \text{ A/cm}^2$ and $dI_{RG}/dt = -500 \text{ A}/\mu\text{s}$ at 125°C . The results of the snubberless SOA determination are shown in Table II. During snubberless measurements both the CGTO and reference devices reached the limit of their static breakdown voltage without destruction [6]. In Fig. 3 the turn-off losses per pulse are shown for the CGTO and reference devices at a current density of 110 A/cm^2 . The reference data are after three consecutive electron irradiation. The active areas of the CGTO and reference devices are 1.8 and 2.3 cm^2 , respectively.

4. Simulations

The simulation effort was focused on the analysis of the static parameters I_{GT} and V_T . The simulations were done using MEDICI (TMA). In the first place, it is well known from experience that both gate trigger current and on-state voltage are easily increased by the reduction of emitter dimensions. The reasons for this are both the reduction of the total emitter area and the increase in the total length of the emitter-gate boundary. The total cathode emitter area is only about 15% of the CGTO device area. The corresponding number for the reference devices is about 50%. The total length of the emitter junction periphery is about two times larger in the CGTO devices compared to the reference devices. Secondly, considering the sensitivity of these two parameters to the surface passivation, care should be taken to ensure the quality of the surface after RIE processing. The first devices processed successfully displayed some peculiarities with respect to the I_{GT} and V_T . First of all, devices showed a distinct turn-on only above $\sim 27^\circ\text{C}$. Secondly the on-state voltage measured at 80 A/cm^2 and 220 A/cm^2 is higher than simulated values assuming an ideal surface ($s = 0$).

In Fig. 4 the simulated $I(V)$ characteristics are shown together with measured on-state voltage values at anode currents of 150 A and 400 A , respectively, for $390 \mu\text{m}$ and $550 \mu\text{m}$ thick CGTO devices. The simulated $I(V)$ curves are adjusted by the choice of the surface recombination velocity, s . The best fit is found for $s = 1500 \text{ cm/s}$. In the case of the $390 \mu\text{m}$ devices two kinds of $I(V)$ curves were observed. Devices displaying higher on-state voltage values (A) have been identified as having an expected final emitter depth of $2.5 \mu\text{m}$. The lower on-state voltage value displayed by some of the devices (B) can be explained by assuming a $0.5 \mu\text{m}$ larger emitter depth, as demonstrated in Fig. 4. This can reasonably be related to the variations in the POCl_3 process. The high-level injection carrier lifetime is $20 \mu\text{s}$ and $40 \mu\text{s}$ for $390 \mu\text{m}$ and $550 \mu\text{m}$ thick devices respectively, and was measured by the contactless optical double laser pump-probe method [7, 8]. The doping profiles were obtained from SUPREM – 3 (TMA) process simulations checked against the resistivity, thickness and profile data obtained from the process monitors. In Fig. 5 simulated I_{GT} versus junction temperature data are shown for $550 \mu\text{m}$ thick devices (C) and compared to the measured results. The models used in the simulations include bandgap narrowing, surface recombination velocity, and temperature and doping dependent lifetime. The temperature dependence of the lifetime is modeled according to $\tau = \tau_0(T/300)^n$ where n is set to 1.65 for holes and 1.77 for electrons [9]. The doping depen-

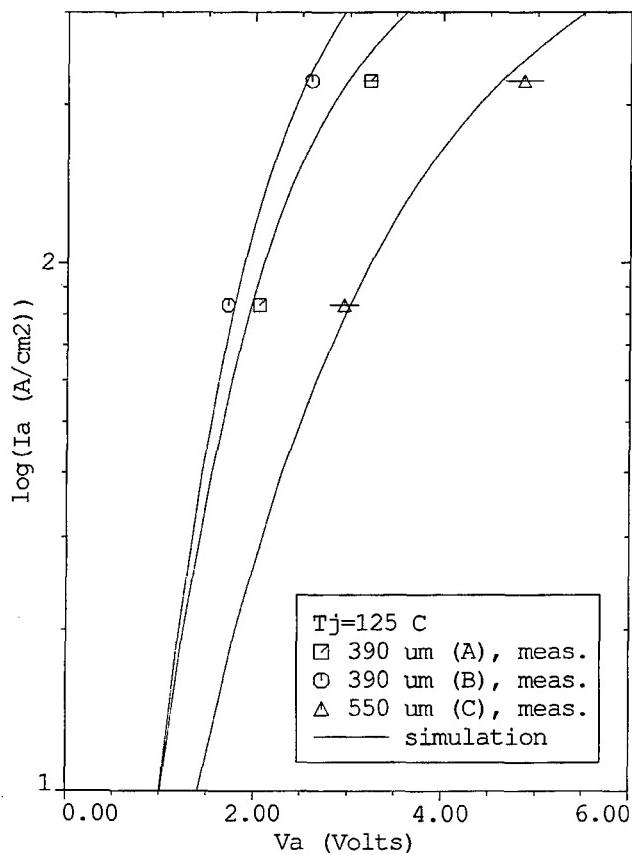


Fig. 4. Simulated (solid lines) and measured (data points) values of on-state voltage for the $390 \mu\text{m}$ (type A and B) and $550 \mu\text{m}$ (type C) thick CGTO devices. The simulated $I(V)$ curves are for surface recombination velocity $s = 1500 \text{ cm/s}$.

dence of the lifetime is modeled by setting $\text{NSRHN} = 9 \cdot 10^{17} \text{ cm}^{-3}$ for electron lifetime, τ_n , and $\text{NSRHP} = 3 \cdot 10^{17} \text{ cm}^{-3}$ for hole lifetime, τ_p [9, 10].

5. Discussion

To check the agreement between simulated and measured values, we adopted the following procedure. First of all, the

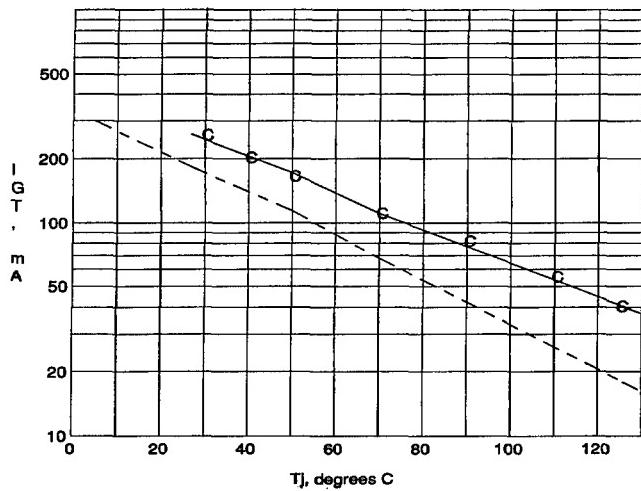


Fig. 5. Simulated (lines) and measured (data points) values of gate trigger current versus temperature for type C CGTO devices. The simulated curves are shown for two values of surface recombination velocity $s = 1500 \text{ cm/s}$ (the solid line), and $s = 0$ (the dashed line). The simulated values shown only in the temperature range where the turn-on is distinct and well defined.

emitter depth and doping level were determined using profiling, junction delineation techniques, resistivity data and process simulation. Secondly, the value of s was adjusted for the best agreement between simulated and measured on-state voltage values. Finally, the I_{GT} versus junction temperature values were simulated with the s value determined from the V_T simulations.

The CGTO devices show high on-state and gate trigger current values even for high values of carrier lifetime in as-processed devices ($\tau = 20\ \mu\text{s}$ and $\tau = 40\ \mu\text{s}$ for $390\ \mu\text{m}$ and $550\ \mu\text{m}$ thick devices, respectively). See also Table I and Fig. 3. Simulations confirm the increase in the on-state voltage with reduced GTO unit cell dimensions. This increase in forward voltage drop has normally to be compensated by increased carrier lifetime. The penalty is however increased turn-off losses as can be seen in Fig. 3. Securing a reasonable low level of on-state voltage and turn-off losses is of great importance for the realisation of fine pattern devices in general. The CGTO devices could only reach the trade-off between on-state voltage and turn-off losses comparable to reference devices when proton irradiation was used [11].

The measured on-state voltage values are higher than the ones calculated assuming an ideal surface ($s = 0$). We identified the critical processing steps responsible for this discrepancy using process and device simulation and comparing the results with experimental process monitoring and device evaluation data. The original CGTO process ($390\ \mu\text{m}$ thick devices) uses relatively shallow, $3 \pm 0.5\ \mu\text{m}$ deep, emitters (emitter depth after POC1_3 process). The wet oxidation results in $\sim 1.7\ \mu\text{m}$ thick oxide, which is necessary to provide a mask during the RIE mesa etch. This leads, however, to a reduction of the emitter thickness. Following the RIE mesa etch, a 30 second $95:5$, $\text{HNO}_3 : \text{HF}$, wet silicon etch is carried out to remove the surface damage prior to dry thermal oxidation. Single oxidation (dry $0.6\ \mu\text{m}$ thick oxide) is carried out, in the case of the $390\ \mu\text{m}$ thick devices, providing the passivation over the cathode junction. Two separate thermal oxidations, the spacer oxidation (dry $0.6\ \mu\text{m}$ thick oxide) and passivation oxidation (dry $0.6\ \mu\text{m}$ thick oxide), are carried out in the case of the $550\ \mu\text{m}$ thick devices. The emitter depth is further reduced by these oxidations. In simulations the reduced emitter depth and nonideal surface are clearly seen to increase the V_T and I_{GT} values. In the $550\ \mu\text{m}$ thick devices the POC1_3 process was modified to give a deeper cathode emitter. The emitter depth in this case is $4.5 \pm 0.5\ \mu\text{m}$ after the POC1_3 process. However due to the emitter thickness reduction by oxidation, the final emitter junction depth was about $2.5\ \mu\text{m}$ in the case of both the 390 and $550\ \mu\text{m}$ thick devices.

An interesting feature in the temperature dependence of the I_{GT} is that the CGTO devices do not trigger below a junction temperature of approximately 27°C . They display a sluggish transistor like behaviour and do not display a distinct latch-up at gate current levels up to 2 A , which is the limit of our equipment. Simulations show that the threshold temperature at which the turn-on behaviour becomes sluggish is influenced by the value of s . This threshold temperature obtained from simulations was $\sim 5^\circ\text{C}$ and

$\sim 27^\circ\text{C}$ for $s = 0$ and $s = 1500\ \text{cm/s}$ respectively, as shown in Fig. 5. To reproduce the temperature dependence of the I_{GT} it is essential to correctly model other factors like temperature dependence of the lifetime and bandgap narrowing. Considering the complexity of the problem, however, the I_{GT} level, the temperature dependence and the threshold temperature are all reproduced reasonably well (see Fig. 5) using the value of $s = 1500\ \text{cm/s}$ and MEDICI standard models for other physical quantities if not stated otherwise.

6. Conclusions

The first fine-pattern CGTO devices fabricated with a newly developed self-aligned process have been shown to have a dynamic performance equal to or better than the standard devices. The on-state voltage and gate trigger current of the CGTO devices are however poorer than the standard GTO devices. The reasons for that, in addition to those related to the unit cell dimensions, have been identified to be a shallow final cathode emitter depth and low quality of the silicon interface after passivation. The present process sequence should be modified at two points. Firstly, the original depth of the cathode emitter should be at least $7\ \mu\text{m}$ preceding the oxidation and mesa etch. Secondly, the surface treatment following the RIE and preceding the thermal oxidation should be improved in order to produce an interface with a low concentration of surface states. The importance of the silicon-silicon dioxide interface quality to the on-state voltage and gate trigger current of the GTO devices have been demonstrated here for the first time. Finally, a good agreement has been achieved between simulated and measured V_T and I_{GT} values using the surface recombination velocity $s = 1500\ \text{cm/s}$.

Acknowledgements

Most valuable contributions from M. Ljungberg, in the first part of this work, are gratefully acknowledged. Thanks also to ABB Drives, Västerås and ABB Semiconductor AG, Lenzburg, for their assistance with the edge treatment. This work was funded by the Swedish IT-4 program (ABB/NUTEK).

References

1. Ishidoh, M., Tada, A., Tokunoh, F., Yamamoto, M. and Nakagawa, T., *JEE Power Electron. and Variable Speed Drives Proc.*, (1990).
2. Satou, Y., Yatsuo, T. and Sakurada, S., *IEEE Trans. on Electron Dev.* **37**, 2034 (1990).
3. Takahashi, Y. *et al.*, *EPE-MADEP Conf. Proc.*, (1991).
4. Bakowski, M., Ljungberg, M., Elderstig, H. and Norlin, P., *1992 Int. Symp. Power Semicon. Dev. Proc.*, *ISPSD'92*, 164 (1992).
5. Elderstig, H. and Kaplan, W., *15th Nord. Semicon. Meeting Proc.*, 15 (1992).
6. Wirth, W. F., *IEEE Trans. Indust. Appl.* **24**, 127 (1988).
7. Waldmeyer, J., *J. Appl. Phys.* **63**, 1977 (1988).
8. Grivickas, V., Linnros, J., Vigelis, A., Seckus, J. and Tellefsen, J. A., *Solid State Electron.* **35**, 299 (1992).
9. Bleichner, H., private communication.
10. TMA Medici User Manual, vers 1.1, 1, 2–3 (1993).
11. Hallér, A., Bakowski, M. and Lundqvist, M., *Solid State Electron.* **36**, 133 (1993).

Application of III-V-Semiconductor Based Heterojunction Bipolar Transistors Towards Multi-Gbit/s 4 : 1 Multiplexer

G. Schuppener, B. Willén, M. Mokhtari and H. Tenhunen

Royal Institute of Technology, Electronic System Design Laboratory, KTH-Electrum, Box 229, 16440 Kista, Sweden

Received May 2, 1994; accepted June 15, 1994

Abstract

An HBT-based IC prototyping technology is being developed at The Royal Institute of Technology's Semiconductor Laboratory, aiming for transistor cut-off frequencies in the region of 70 GHz. Both AlGaAs/GaAs- and InP/InGaAs-HBTs have been fabricated showing cut-off- and maximum oscillation frequencies well above 60 GHz. The overall circuit technology is based on two level metallization separated by silicon-nitride for interconnects and MIM capacitors, and NiCr-based resistors.

A 4 : 1 time division multiplexer was targeted to explore the feasibility of the in-house technology for MSI circuit implementations, especially for broadband switching applications.

1. Introduction

If the material within which the emitter of a bipolar junction (BJT) is defined has a larger bandgap than the material for the base in the same transistor, the emitter-base junction is said to be a heterojunction and the device is named Heterojunction Bipolar Transistor (HBT). The HBT still have the advantages of a silicon bipolar transistor, but extend them to higher frequencies [1, 2].

Due to the different bandgaps at the emitter-base heterojunction an energy barrier ΔE_g is created at the interface. A spike, which appears in the conduction band, tends to retard the flow of electrons from emitter to base. At the first glance, this seems to be a disadvantage, as the emitter injection efficiency is reduced. But it provides benefits for high-speed operation, because electrons that surmount the energy barrier are injected into the base with high forward velocities, decreasing the base transit time. Furthermore, due to the step in the valence band, the flow of holes from base into the emitter region is suppressed, increasing the current gain in turn.

It is then possible to ensure sufficient current gain with high levels of base doping N_{ab} , even higher than the emitter doping level N_{de} , as can be seen from eq. (1), where D_{nb} denotes the diffusion constant of electrons in the base, W_b the base width, D_{pe} the diffusion constant and L_{pe} the diffusion length of holes in the emitter and β_{BJT} the current gain of a standard homojunction bipolar transistor. This high base doping concentration level offered by the HBT leads to low base sheet resistance and emitter junction capacitance and high transistor f_T and f_{max} [eq. (2) and eq. (3)].

$$\beta_{HBT} = \frac{D_{nb} L_{pe} N_{de}}{D_{pe} W_b N_{ab}} \cdot \exp\left(\frac{\Delta E_g}{kT}\right) = \beta_{BJT} \cdot \exp\left(\frac{\Delta E_g}{kT}\right) \quad (1)$$

Further, the used III-V semiconductor materials, especially InGaAs, have high electron mobility, which improves the high frequency characteristics. Also access to semi-insulating

(SI) substrate materials makes the parasitics very low and improves the high frequency aspects.

2. Comparison of the InP/InGaAs and the AlGaAs/GaAs system

As already mentioned, one major advantage of InGaAs is the high electron mobility, which can be 1.6 times higher than for GaAs and 9 times higher than for Si. In addition, the extent of transient velocity overshoot is greater than in GaAs, resulting in lower transit times. The bandgap of InGaAs is smaller than that for GaAs or Si, which is directly related to the V_{be} -threshold voltage. Therefore, supply voltage and power dissipation can be decreased, which offers larger complexity and improves the power-delay product, a figure of merit in logic circuit design. Thermal conductivity, important for heat-sinking, is higher for InP substrates, but still below that for silicon.

Aluminium shows a high affinity to oxygen, which especially at a heterojunction interface can degrade electrical properties by reducing mobility and enhancing carrier trapping. Due to the Al content AlGaAs suffers from high recombination velocity at the surface. The recombination velocity is smaller in InGaAs by a factor of 1000. For HBTs this effect becomes important, when the device is scaled down to small dimensions and the emitter area-to-periphery ratio will decrease. As a result, there is less base current in a InP/InGaAs-HBT [1, 2].

InP-based semiconductor lasers have the fastest modulation capabilities and the optical devices are operating in the 1.3–1.55 μm wavelength range, where silica-based fibers show low dispersion and low losses. Therefore an HBT technology in the InP-system would be an excellent base for future opto-electronic ICs.

A major disadvantage of the InP-based system, influencing both fabrication and performance, is the fact that it is difficult to produce high-resistivity regions by implantation in InP, and impossible to do so in InGaAs. Furthermore, it is necessary to grow the epitaxial structure with fixed alloy compositions in order to achieve lattice matching. The requirements for the growth process are increased and grading of alloy compositions, widely used in AlGaAs/GaAs-HBTs, is not possible.

3. Device fabrication

In the case of the AlGaAs/GaAs-HBT the epitaxial structure is grown by molecular beam epitaxy (MBE) on semi-insulating GaAs substrate. A linearly graded aluminium

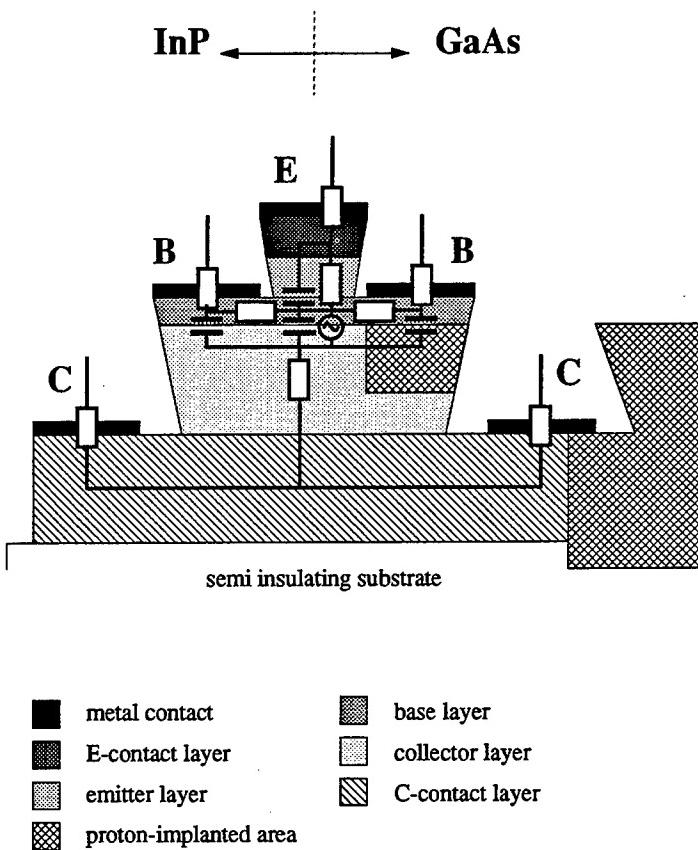


Fig. 1. Device cross section and small-signal equivalent circuit related to the device structure.

concentration in the base is used to achieve a built-in electric field and improve electron transport through the base. Standard optical photolithography and lift-off techniques are used in the fabrication process. The emitter contact with an area of $2.5 \times 10 \mu\text{m}^2$ also acts as a mask while etching down to the base layer, defining areas for base contacts and for a shallow proton implant into the collector layer, which reduces the extrinsic base-collector capacitance. To minimize the undercut of the emitter mesa, a combination of both dry and wet etching is used. Si_3N_4 -side-walls are employed to define the spacing to the self-aligned base contact. A base to emitter distance of about $0.2 \mu\text{m}$ is achievable. Finally the collector is connected through a non-self-aligned via down to the highly doped subcollector. Si_3N_4 deposited by plasma enhanced chemical vapor deposition (PECVD) is used as dielectric and evaporated AuGe/Ni/Ti/Au and Au/Ni/Ti/Pt/Au for n- and p-type contacts, respectively. All ohmic contacts are alloyed at the same time at 360°C . Device isolation is achieved by deep implantation of protons. Sputtered TiW/Au and NiCr are used for two level circuit connections and resistances (sheet resistance $50 \Omega/\text{square}$), respectively. The temperature during the process never exceeds 360°C in order to avoid damage of the structure.

By modifying some of the processing steps the InP/InGaAs-HBT can be fabricated using the same mask set as for the GaAs-based device. Here, the epitaxial structure is grown on an SI-InP substrate. The possibility of producing semi-insulating regions using proton implantation does not exist for the InP/InGaAs-HBT, where isolation is carried out by etching the epitaxial structure down to the SI-substrate. The active collector area can not be diminished,

resulting in a large extrinsic capacitance. In spite of that unfavourable ratio of intrinsic to extrinsic parasitics, the device shows a high-frequency behaviour as good as the AlGaAs/GaAs-HBT and it is apparent that there is considerable interest in reducing the active area. Furthermore, as a result of the mesa-structure and the deep vias to the collector and the SI-substrate, the surface becomes very non-planar and there can be concerns about the step coverage of metallizations for InP-based ICs.

A cross section of the resulting mesa-structure for both the AlGaAs/GaAs- and the InP/InGaAs-HBT is given in Fig. 1, showing also the physical background for extraction of the small-signal equivalent circuit, presented in the next paragraph.

An important issue of the fabrication process is the self-aligned base contact to decrease parasitics. Various process flows have been reported with somewhat different self-alignment techniques, but comparable high-frequency performance of the device [3, 4].

4. Results

The microwave response of the transistors was evaluated using an automatic network analyzer for frequencies up to

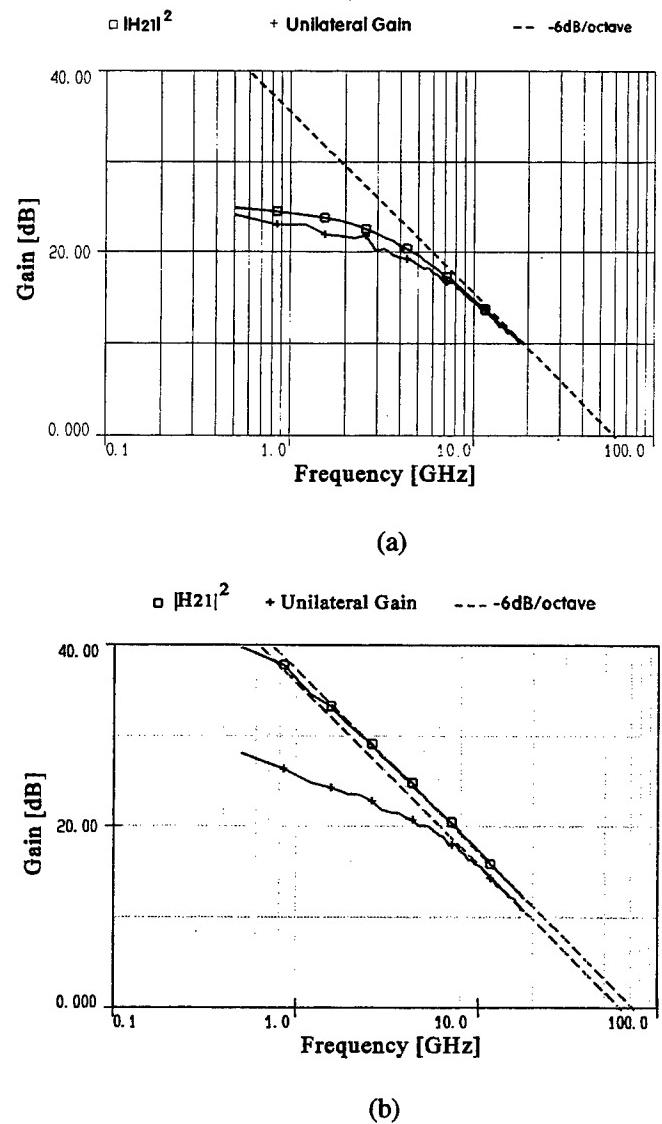


Fig. 2. High frequency gain curves for an AlGaAs/GaAs- (a) and an InP/InGaAs-device (b).

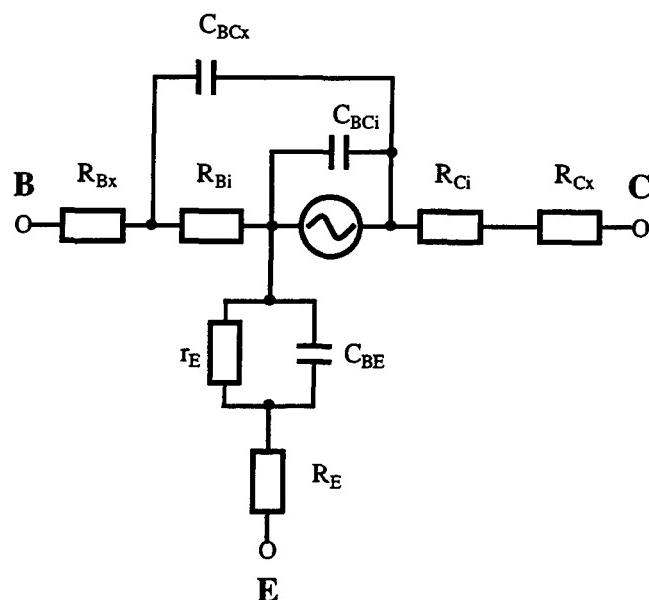


Fig. 3. Small-signal equivalent circuit.

18 GHz. The measured S-parameters were then used to calculate the cut-off frequency and the maximum-oscillation frequency by extrapolating the measured data by $-6 \text{ dB}/\text{Oct}$. The high-frequency gain curves are shown in Fig. 2 for an AlGaAs/GaAs- and an InP/InGaAs-device.

From these measurements the parameters of the small-signal equivalent circuit, shown in Fig. 3, were extracted with eqs (2) and (3) as basis of the evaluation. The results are presented in Table I, compared with a fast silicon homojunction bipolar device, fabricated at the same laboratory.

$$\frac{1}{2\pi f_T} = r_E C_{BE} + (r_E + R_E + R_C) C_{BC} + \tau \quad (2)$$

$$f_{\max} = \sqrt{\frac{f_T}{8\pi R_B C_{BC}}} \quad (3)$$

5. Circuit demonstrator

Till now the technology described above was used to fabricate single devices or SSI circuits up to 100 devices like NTL ring oscillators. To evaluate the technology on MSI circuit design level, a 4 : 1 multiplexer was designed with

respect to the AlGaAs/GaAs-system first. The aim is to have the possibility of prototyping high-speed and optoelectronic ICs.

Time division multiplexers operating at gigabit per second (Gbit/s) output rates are required for digital optical communication systems in order to exploit the high transmission capacity of single-mode optical fibers. At the transmitting-side of a communication system, multiplexers are combining the incoming parallel data channels to one single data stream with high bit rate. The schematic of the 4 : 1 multiplexer is shown in Fig. 4.

Differential operation, the use of both inverted and non-inverted signals, was applied for the internal as well as for the output signals. This gives the possibility to reduce the logic voltage swing, resulting in faster circuit operation, and minimizes crosstalk between wires. The data inputs are single-ended to ensure ECL compatibility.

For multiplexing a two stage architecture consisting of three 2 : 1 MUXs has been chosen. Thereby, only one MUX is operating at the maximum bit rate, whereas the other MUX-stage is operating at half this rate. Therefore, the incoming system clock has to be divided by two by a static frequency divider, which also generates the different phases for clocking the input latches.

To switch the second stage MUX, a high-speed signal with the same frequency as the system clock is required. This signal is generated by a XOR-gate, which acts as frequency doubler when driven by two equal frequencies exactly 90 degrees out of phase. This has the advantage that the XOR-gate places the edges of the doubled frequency one propagation delay away from the edges of the original frequencies. Hereby, the dynamic delay for data to pass through the first MUX stage is compensated.

The input data is latched parallel and is piped through the three stage input flip-flops by one fourth the maximum bitrate. The use of the pipelined input increases the maximum bit rate considerably. The clock phases are distributed to the D-type Master-Slave Flip-Flops (DMSFFs) and the MUXs in such a way that no input of a flip-flop is clocked closer than two bit duration time after the output of the previous stage, as well as there is the same time margin between the last DMSFF stage and the first MUX stage.

The same gate topology is used to create the XOR, MUX and latches, in order to preserve device matching and operation uniformity. As a result, the delay times for all cells are nearly the same and the propagation delay of the basic logic

Table I. Small-signal parameters

Variable	Si	GaAs	InP	Unit
r_E	2.8	1.6	1.2	Ω
R_E	3.8	5.7	7.6	Ω
R_B	50	33	34	Ω
R_{Bx}	60	10	9.1	Ω
R_C	29	5	11	Ω
C_{BE}	1060	116	148	fF
C_{BC}	16	11	14	fF
C_{BCx}	18	18	11	fF
τ	7.0	2.1	1.5	ps
f_T	15	63	73	GHz
f_{\max}	12	65	62	GHz
β	100	18	164	

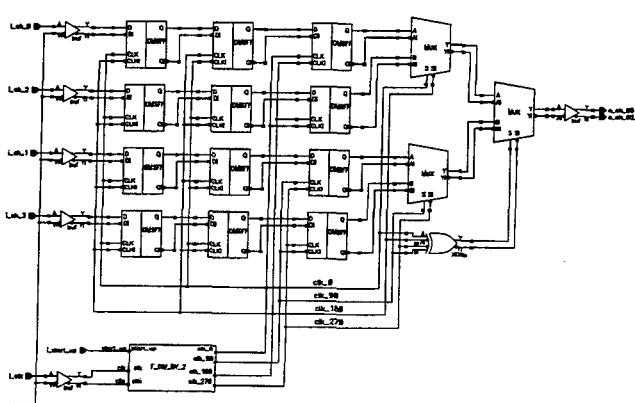


Fig. 4. Block diagram of the 4 : 1 multiplexer.

cell is the only parameter restricting the maximum theoretical bit rate [5, 6].

6. Simulation and layout

Simulations were done for all single device building blocks and for the complete multiplexer including wiring parasitics at operating speed sufficient for 4 Gbit/s data rate. For HSPICE-simulation the standard Gummel-Poon transistor model was applied to describe the HBTs. The Spice model parameters, shown in Table II, were extracted from measurements of older devices with lower f_T . These devices were fabricated on AlGaAs/GaAs MOCVD-grown epitaxial structures [7].

Fig. 5 shows the simulation diagram of the 4 : 1 multiplexer with clock, input-channels 0 and 1 in antiphase, channels 2 and 3 tied to fixed logic values, and the differential output signal at 4 Gbit/s. Nevertheless, simulation results have to be interpreted with care. The used Gummel-Poon transistor model is not able to consider certain effects specific for the HBT, especially when operating at the frequency limit. Furthermore, the HF-parasitics have not been extracted for every signal path. But the simulation gives a good reliability for circuit operation and will be the base for future model optimization.

As already mentioned, the circuit is realized for feasibility studies of the in-house technology with respect to MSI circuit implementations. Therefore, the layout was optimized for high yield, and for verification of the HF-parasitics estimations done during the simulations. Wiring parasitics and cross-coupling caused by the layout structure

Table II. Spice model parameters

AlGaAs/GaAs-HBT		
IS = 3E-24	BF = 18	NF = 1.1
RE = 25	RC = 40	RB = 30
CJE = 120F	VJE = 2.3	MJE = 0.4
CJC = 40F	VJC = 1.4	MJC = 0.5
CJS = 7F	VJS = 1000	MJS = 0.5
FC = 0.5	NE = 5	TF = 6P

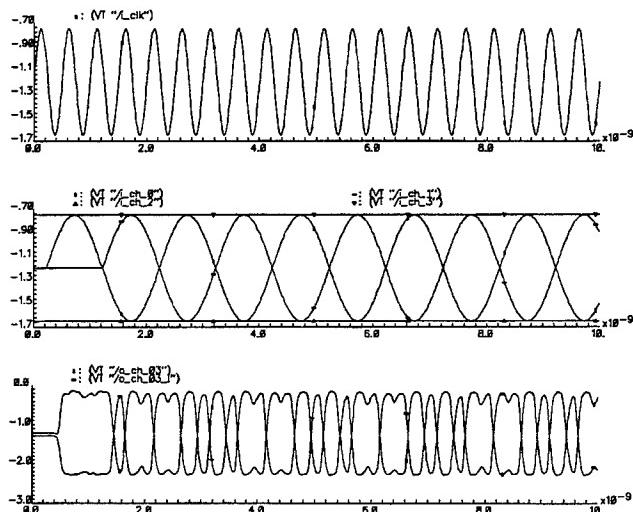


Fig. 5. Simulation of the 4 : 1 multiplexer at 4 Gbit/s output rate.

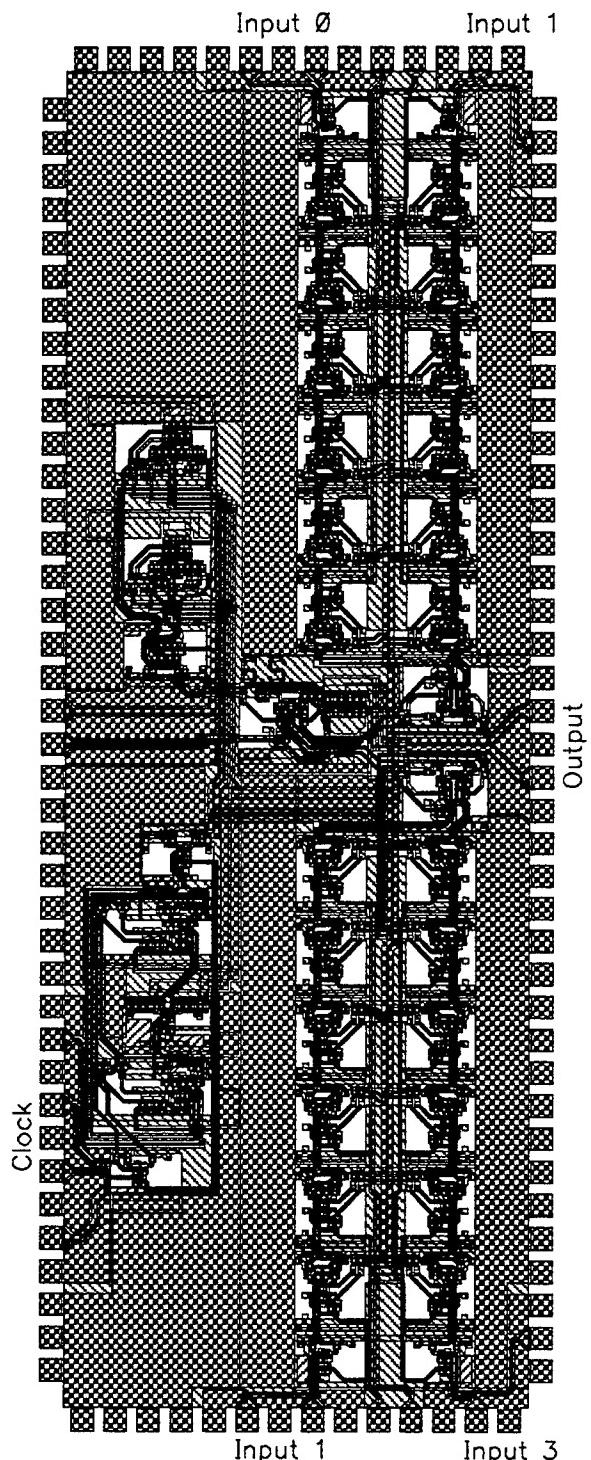


Fig. 6. Layout of the 4 : 1 multiplexer.

shown in Fig. 6 were extracted and simulated using inhouse software [8, 9]. The wires are modeled as LRC-based filters. The distributed network caused by wiring is extracted from the layout by modelling the discontinuities resulted from wiring redirections as series resistance and series inductance between nodes and capacitive coupling to ground, scaled by the dimensions of the actual current terminal (wire).

Cross coupling is modelled as capacitive coupling between parallel wires, scaled as a function of the distance. This was done only on frequency-critical wires in the circuit. The same software was used to extract performance of the ring oscillators, mentioned above as the first circuits realized in that technology. Simulation of these circuits resulted in a

delay time of $\tau_d = 35$ ps with a power consumption per gate of $P_{W/gate} = 70$ mW, whereas the measured results were only slightly different with $\tau_d = 34$ ps and $P_{W/gate} = 65$ mW.

7. Conclusion

We have fabricated InP- and GaAs-based HBTs showing cut-off frequencies well above 60 GHz and sufficient dc current gain for IC design. Both technologies are quite well developed and performance is comparable with international foundries.

Using this in-house technology we have designed a 4:1 multiplexer for operation at high output rates. Currently the circuit is under fabrication at the Semiconductor Laboratory of the Royal Institute of Technology, and measurements are expected soon. The circuit area is 2.3×6.3 mm² and the power dissipation is expected to be approximately 18 W, including the output load.

Further circuit optimization with respect to better device models can be expected to greatly improve the performance, both on higher speed and lower power consumption.

Acknowledgements

We gratefully acknowledge Dan Haga for device processing, Dr Urban Westergren for performing the measurements and Hans Berggren for

setting up the computer system. This work is financed by NUTEK under contracts 93-02955P and PS 93-2962 DS.

References

1. Sze, S.M. (Ed.), "High-Speed Semiconductor Devices" (John Wiley & Sons Inc. 1990)
2. Asbeck, P. M., Chang, F. M.-C., Wang, K.-C., Sullivan, G. J. and Cheung, D. T., "GaAs-Based Heterojunction Bipolar Transistors for Very High Performance Electronic Circuits", Proceedings of the IEEE **81**, 1709 (1993).
3. Chang, M.-C. F. et al., "AlGaAs/GaAs Heterojunction Bipolar Transistor Fabricated Using Fully Self Aligned Dual Lift-Off Process", IEEE Electron Device Letters **EDL-8**, 303 (1987).
4. Nittono, T., Nagata, K., Nakajima, O. and Ishibashi, T., "A New Self-Aligned AlGaAs/GaAs HBT Based on Refractory Emitter and Base Electrodes", IEEE Electron Device Letters **EDL-10**, 506 (1989).
5. Negus, K. J., "Multi-Gbit/s Silicon Bipolar Multiplexer and Demultiplexer with Interleaved Architectures", Proceedings IEEE Bipolar Circuits and Technology Meeting, pp. 35–38, 1991.
6. Reimann, R. and Rein, H.-M., "A 4:1 Time-Division Multiplexer IC for Bit Rates up to 6 Gbit/s Based on a Standard Bipolar Technology", IEEE Journal of Solid-State Circuits, **SC-21**, 785 (1986).
7. Willén, B., Mokhtari, M. and Westergren, U., "III-V-Semiconductor based Heterojunction Bipolar Transistors (HBT)", Swedish Institute of Microelectronics, 1992.
8. Lundén, H., "Extraction of Distributed Networks in IC Design", Swedish Institute of Microelectronics, Report IM I 88 3036.
9. Lundén, H., "Parasitics in IC Interconnects on GaAs", Swedish Institute of Microelectronics, Report IM I 88 3036.

Multichamber Processor for Small Semiconductor Laboratories – the First Results

R. Punkkinen, H. Ihantola, K. Jokinen, T. Kuusela, H. Arvela, H-P. Hedman and V. Saikku

Laboratory of Electronics and Information Technology, Department of Applied Physics, University of Turku, Lemminkäisenkatu 14 A, FIN-20520 Turku, Finland

Received May 9, 1994; accepted in revised form July 5, 1994

Abstract

At the University of Turku in the Laboratory of Electronics and Information Technology a microcomputer controlled compact multichamber processor for semiconductor wafer processing has been designed and built. Several different processes like various chemical vapor depositions (CVD), plasma etching and sputtering can be carried out. At the moment the research processor contains two reactors for 150 mm silicon wafers and a third one for sputtering is under construction. The dimensions of the processor including gas cabinets and UHV pumping equipment are about ($L \times H \times W$) $4 \times 1.2 \times 1 \text{ m}^3$. The objectives of this system are to minimize the need of clean room space without compromising the clean process conditions and to provide the means for laboratory scale integrated circuit fabrication. For now this system can be used for processing CVD oxidation, both epitaxial and polysilicon growth and plasma etching. In the CVD reactor we have grown e.g. silicon homoepitaxial layers of good quality at low pressure and temperature (0.5–200 Pa, 650–750 °C) at the growth rate of about 5 nm/min.

1. Introduction

Semiconductor technology has evolved fast during the last three decades. New composite materials like gallium-arsenide have been developed, although silicon will still be the most common semiconductor. The lateral element dimensions have been reduced below 1 μm and the size of wafers have grown to 25 cm. However, the use of large size wafers has raised the questions of handling, processing position and thermal stress.

The prices of semiconductor processing equipment have always been high. When the requirements of the accuracy and the stability of temperatures and other processing parameters have been tightened, the prices have increased, and now they exceed the resources of most universities. At universities and in many research institutes there is no need for volume production, and it is enough, if one wafer at a time is to be processed and carried sequentially through various steps [1]. Thus, it is useful to build a versatile central unit, common to all different reactors, equipped with power supplies, automatic controls, vacuum pumps, and gas distribution, which will serve several different reactors. The reactors would be used only one in turn. The required space for the equipment and the operation area can be reduced below 30 m². The actual footprint of the device here is only about 4 m². The lithography equipment and optical as well as electrical measuring instruments will be needed in addition.

2. Description of the system

At the University of Turku in the Laboratory of Electronics and Information Technology we have been designing and constructing a research multichamber processor for the 150 mm silicon wafers since 1991. The goal has been to

build a modern compact silicon semiconductor processing equipment which should need only a small clean room space and form a complete system for various processes e.g. towards novel structures of silicon IC's. At the moment the equipment consists of three typical reactors with a combined vacuum system (Fig. 1). The vacuum system consists of a common element, a turbo pump (880 l/s), a mechanical pump (63 m³/h), and HV/UHV valves. The reactors are connected via flexible tubing and vacuum valves to the expandable common element which can also be used to collect residual gases for spectrometric analysis. The dimensions of the processor including gas cabinets are about ($L \times H \times W$) $4 \times 1.2 \times 1 \text{ m}^3$. The first reactors are aimed for chemical vapor deposition, etching and sputtering.

At the moment, the gas distribution system contains high purity process gases (Ar, H₂, N₂, O₂, AsH₃, B₂H₆, SiH₄, CCl₄ and CF₄) of which hazardous gases are placed in gas cabinets locating next to the reactors. The gas piping is made of 1/4" ultrapure electropolished stainless steel (AISI 316L) tubes. The gas flows are controlled and regulated by computer operated pneumatic valves and mass flow controllers. The high purity of gases near the point of use is secured with particle filters. The complete gas and vacuum systems are leak tested with helium.

Environmental aspects and low gas consumption were taken into account. Thus it is possible to keep the amount of doping gases very low. Even in the event of a severe accident, the concentration of the doping gases in room air would stay below 0.3 ppm. In addition, the system is equipped with a hazardous gas alarm system. The exhaust gases from the pumps can be fed through a microprocessor controlled combustion unit where propane-air flame burns all exhaust gases in a ceramic tube at the temperature of 2000 °C.

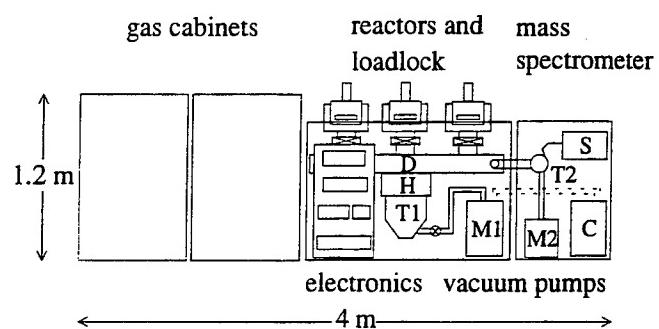


Fig. 1. The schematic layout of the Multichamber processor. The main parts of the processor are a common vacuum element D, mechanical pumps M1 and M2, turbo pumps T1 and T2, a throttle valve H, a mass spectrometer S, and a combustion unit C including reactors and gas cabinets. The width of the system is about 1 m.

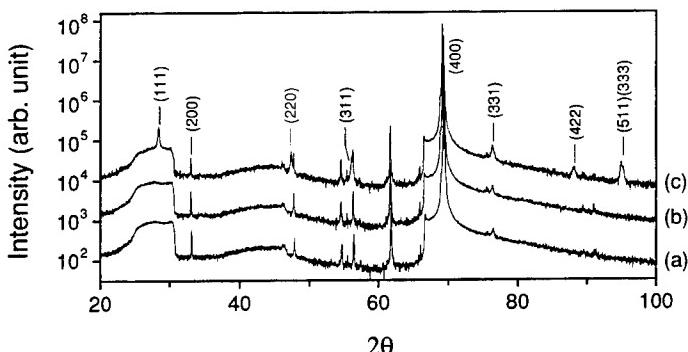


Fig. 2. The X-ray diffraction intensities as the function of the measurement angle 2θ using $\text{CuK}_{\alpha 1}$ -radiation. The intensity-scale refers to the curve a, the curve b and c are shifted one and two decades, respectively. The curve a is from the reference measurement of an unprocessed wafer, the curve b is from a typical processed wafer with a 2- μm epitaxial layer and the curve c is from a wafer with a 6- μm thick polysilicon layer grown on it. The Miller indices show the allowed reflections from the silicon crystal. The strong (400) reflection peaks come from the (100)-oriented monocrystalline substrates.

At the moment there are two routinely operative reactors in the processor, the first one for CVD processes and the second one for plasma etching. The third one for sputtering is now under construction. The whole system is regulated and controlled by programmable logic units operated by a PC.

Some semiconductor process steps cannot be executed in the multichamber processor and have to be done outside. For example, lithographic operations are processed separately. Also all the necessary wet cleaning and etching steps require other facilities. For these processes there is another small clean room for safe handling of the processed wafers. The transportation of the wafers between the work stations is done using special cassettes compatible with the loadlock of the processor.

3. Reactor for CVD processes

The inside measures of the first process chambers (CVD and etching) are 320 mm in diameter and 300 mm in height and they are made of stainless steel (AISI 304). The chambers aimed for high temperature processes are water cooled. The

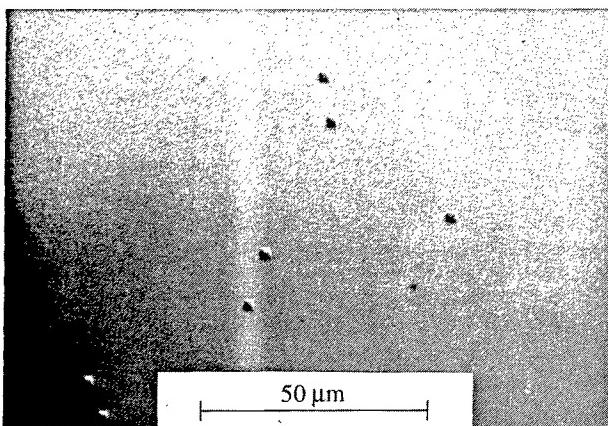


Fig. 3. A photograph using Nomarski microscopy of the defect etched 1- μm epitaxial layer. Yang-etch [6] was used for 180 s. The wafer was processed before the ambient air on the wet cleaning station was properly filtered and thus there exists stacking faults.

first reactor was designed for epitaxial and CVD processing with a separate loadlock connected to the mechanical pump to ensure clean processing conditions. The gate valve opening between the loadlock and the reactor is 210 mm \times 25 mm. The processing temperature of the wafer is planned to be below 900 °C. In the beginning, an electric heating method was selected. The dimensions of the graphite heater are 200 \times 200 \times 5 mm³ and the maximum heating power is 3.6 kW. The wafer being processed is pushed with a steel rod from the loadlock through the gate valve on a 3 mm thick quartz susceptor. Process gases are distributed into the reactor through a special nozzle about 80 mm above the quartz plate. The nozzle consists of four lateral tubes welded together to form a 200 mm \times 200 mm square with 20 small holes (0.2 mm in diameter) through which the process gases are injected.

The growth process for the silicon epitaxy is similar to the UHV/CVD process introduced by Meyerson [2]. Epitaxial layers were deposited on lightly doped n-type (100) silicon wafers with a resistivity of about 1 Ω cm. Before the growth the wafer is dipped into 5% HF solution for 10 to 30 s [3] to remove the native oxide. Then the wafer is immediately moved into the transfer cassette, which is connected to the loadlock. The loadlock is purged with nitrogen and pumped to a prevacuum of about 10 Pa. After that the wafer is pushed into the reactor onto the quartz susceptor under a low pressure nitrogen flow. Before the growth starts the reactor is pumped down below 1 μPa to remove all possible gaseous contaminants. Then the wafer is heated up to the growth temperature of 650–750 °C and process gases are fed into the reactor. The growth pressure has usually been in the range of 0.5–200 Pa. Silicon was deposited from silane SiH_4 with a typical flow rate of 4 sccm. Hydrogen was used as a carrier gas and oxide growth suppressor with a typical flow rate of 20 sccm. Arsine AsH_3 and diborane B_2H_6 were used as doping precursors. Typically, growth rates were found to be about 5 nm/min, which is in agreement with the results published by other researchers [4, 5]. This gives a high growth efficiency of about 5%. At the moment, the possible doping range is 10^{16} – 10^{19} cm^{-3} . CVD SiO_2 films were successfully grown from silane and oxygen. Polysilicon films can also be grown with this reactor. After the CVD oxidation the reactor can be prepared for silicon epitaxy by short nitrogen purge and pumpdown cycles. However, the intense use of the reactor necessitates the periodic cleaning of the chamber.

Consequently, various thin films can be grown on silicon wafer surface. X-ray diffraction measurements showed that the epitaxial layers are of good quality and follow the crystal structure of (100)-oriented substrates. Three examples of the diffraction measurements are shown in Fig. 2 as a function of the measuring angle 2θ using $\text{CuK}_{\alpha 1}$ -radiation. The $\text{CuK}_{\alpha 2}$ -radiation was not filtered and it can be seen in the strongest reflection peaks. The curve a is from a reference measurement of an unprocessed wafer, b is from a typically processed wafer with a 2- μm thick epitaxial layer and the curve c is from a wafer with a 6- μm thick polysilicon layer grown on it. The curves from unprocessed and epitaxial wafers are nearly identical. The curve c from the polysilicon wafer indicates reflection peaks from different silicon crystal planes. The strong (400) reflection comes from the (100)-oriented substrate and it is seen on all curves. The

(200) reflection is basically forbidden in diamond crystal structure but it is seen because of multiple reflections in the crystal.

In order to search for defects the wafer was etched by using the Yang method [6]. Microscopic studies revealed that only occasional stacking faults were found on the surface (Fig. 3). Their number was reduced practically to zero by decreasing the amount of particles locally on the wet clean station.

4. Plasma etching reactor

The plasma reactor is a single wafer parallel-plate system. In this cathode-coupled RIE type reactor the top electrode and the chamber walls are grounded and the lower electrode with the wafer holder is powered through a dc-blocking capacitor. As the ratio of the effective areas of the aluminium electrodes is large, a significantly high sheath dc self-bias potential can be produced. In the typical process pressure range of 5–100 Pa the bias voltage varies from 60 to 800 volts (the frequency of the rf generator is 100 kHz and the power is 65 W.) Therefore the ion bombardment can be adjusted for suitable anisotropy of etch processes [7].

The etching gases are CF_4 and CCl_4 , which can be used to etch most interesting layers, e.g. Si, SiO_2 , polysilicon and Al. Selectivity can be increased with additional gases like O_2 , H_2 , Ar and N_2 [8]. The preliminary results show that the etch rate of Si and SiO_2 is 100 nm/min (pressure 10 Pa, gas flow 30 sccm; etch time 5 min).

In the future the reactor will be equipped with a high frequency generator (13.56 MHz) in order to increase the concentration of reactive radicals and etch rate of various layers.

5. Sputtering reactor

The magnetron sputtering reactor has a water cooled aluminium cathode of 2" in diameter at a distance of about 280 mm from the substrate. A manually operated gate valve will be inserted between the target and substrate, so an easy change of the target material is possible without opening the reactor. The long distance between the target and the substrate enables to use a small and inexpensive magnetron and the deposition will be automatically uniform. The maximum sputtering power of the magnetron is 1 kW. The deposition process can be monitored visually through two windows shielded from deposition by manual shutters. At first only DC sputtering will be used, but in the future the RF sputtering and sputter etching can be carried out, too.

6. Conclusions and future applications

The multichamber processor can be applied to several semiconductor manufacturing processes. In fact, when the pro-

cessor is fully assembled it can perform all the necessary process steps to produce complete integrated circuits excluding lithography and wet cleaning. The benefits with this kind of operation are the cleanliness of the wafer processing in the system and the possibility for low cost manufacturing of small sets of integrated circuits. Also the requirements for the clean room space are reduced and the process gas consumption is extremely low. When the wafers are transported between the work stations the cleanliness is secured by special cassettes.

The CVD reactor can be applied for example for CVD oxide growth with silane and oxygen and for polysilicon growth. Because of the relatively low process temperatures there are no autodoping problems. After the low temperature oxidation just a short nitrogen purge and pump-down cycle is sufficient to clean the reactor ready for epitaxial processing. The epitaxial process itself can be used for several applications including selective epitaxial growth [9, 10] and shallow structures with very sharp doping profiles. The dopant concentration in the growing epitaxial layer can be varied to achieve complex structures. It is also easy to include a germane (GeH_4) gas line to the multichamber processor whereupon it becomes possible to produce variable silicon-germanium structures like quantum wells with the CVD process reactor [11].

The concept of the whole system is completely new and easily extendable to numerous applications so far difficult to attain.

Acknowledgements

The authors would like to thank the Academy of Finland, Okmetic Ltd and Outokumpu Invest Oy for financing this work partially. The role of the University of Turku, which has provided both human and material resources, is gratefully acknowledged.

References

1. Saraswat, K. C. *et al.*, 1989 SPIE Symp. on Microelectronic Processing, 1189-01 (1989).
2. Meyerson, B. S., *Appl. Phys. Lett.* **48**, 797 (1986).
3. Meyerson, B. S., *Appl. Phys. Lett.* **57**, 1034 (1990).
4. Liehr, M., Greenlief, C. M., Kasi, S. R. and Offenberg, M., *Appl. Phys. Lett.* **56**, 629 (1990).
5. Kircher, R., Furuno, M., Murota, J. and Ono, S., *J. Crystal Growth* **115**, 439 (1991).
6. Yang, K. H., *J. Electrochem. Soc.* **131**, 1140 (1984).
7. Horowitz, C. M., *J. Vac. Sci. Technol. A, Vac. Surf. Films* **1**, 60 (1983).
8. Simko, J. P. and Oehrlein, G. S., *J. Electrochem. Soc.* **138**, 2748 (1991).
9. Sabine, K. A. and Kemhadjian, H. A., *IEEE Electron Device Lett.* **6**, 43 (1985).
10. Murota, J., Nakamura, N., Kato, M., Mikoshiba, N. and Ohmi, T., *Appl. Phys. Lett.* **54**, 1007 (1989).
11. Meyerson, B. S., Uram, K. J. and LeGoues, F. K., *Appl. Phys. Lett.* **53**, 2555 (1988).

Design Optimization of Negative Charge Pumps for IGBT driver ICs

Markku H. Åberg

VTT Electronics, Oalarinluoma 9, FIN-02200 Espoo, Finland, PACS no 7390

Received May 2, 1994; accepted in revised form June 29, 1994

Abstract

An optimized design methodology for negative bias charge pumps is proposed. The circuit discussed gives an output voltage negative to the ground and can be used with n-substrate CMOS processes. Negative bias voltages are used in IGBT gate drivers. The analysis in this paper is done for a resistive load, but it can be applied also to capacitive loads in steady state cases with constant average current. The design rules are derived from theoretical calculations and verified with simulations.

1. Introduction

Negative gate biases are used in driving IGBTs for improving the noise margin and preventing dv/dt induced turn-on, and for speeding up the switch-off. The negative gate bias can be supplied from the power supply. However, in many cases it is more economically to generate the negative bias on the driver chip. A circuit called charge pump is generally used for generating this negative bias on chip.

Charge pumps can also be used for generating voltages over the supply voltage (positive charge pump), and some papers have been published about optimizing the design of these [1, 2]. However, the author has found no papers published during the last ten years of the optimized design of negative charge pumps.

2. Theory of negative charge pumps

Basically a negative charge pump consists of two capacitors, pump capacitor C_p and hold capacitor C_h , and four switches for transferring the charge from the power supply via the pump and hold capacitors to the load, Fig. 1.

In the first phase the pump capacitor C_p is loaded from the power supply V_{DD} . In the second phase the charge is transferred to the hold capacitor C_h . If there were no discharge to the load, the output voltage would be after the n th

cycle (1)

$$V_{out}(n) = -V_{DD} \frac{C_p}{C_h} \sum_{i=1}^n \left(\frac{C_h}{C_p + C_h} \right)^i \quad (1)$$

where V_{DD} is the supply voltage. And the limiting value of V_{out} , when $n \rightarrow \infty$ is of course $-V_{DD}$. If the hold capacitor is discharged through a resistive load R_L with a discharging time t_1 , the output voltage after the n th cycle, when omitting the pump capacitor discharge to load time t_2 , is according to eq. (2). For the moment the resistance of the switches, R_{on} , is considered as insignificant. This is true when the time constant $R_{on} C_p \ll t_1, t_2$.

$$V_{out}(n) = \frac{-V_{DD} C_p}{C_h e^{-t_1/(R_L C_h)}} \left(\frac{C_h e^{-t_1/(R_L C_h)}}{C_h + C_p} \right)^n \quad (2)$$

The steady state minimum negative voltage limit, that is the output voltage value after the discharge period t_1 , with resistive load is (when $n \rightarrow \infty$) eq. (3)

$$V_{out}(\infty) = \frac{-V_{DD} C_p \exp(-t_1/R_L C_h)}{C_p + C_h (1 - \exp(-t_1/R_L C_h))} \quad (3)$$

The maximum negative voltage limit is calculated by setting the exponential term after $-V_{DD} C_p$ to unit value. When one takes into account the pump capacitor discharge time t_2 , the equation becomes a little bit more complicated (4)

$$V_{out}(\infty) = \frac{\left(-V_{DD} C_p \exp \frac{-t_1(C_p + C_h) + t_2 C_p}{R_L C_h (C_p + C_h)} \right)}{\left(C_p + C_h \left(1 - \exp \frac{-t_1(C_p + C_h) + t_2 C_p}{R_L C_h (C_p + C_h)} \right) \right)} \quad (4)$$

where V_{DD} is supply voltage, R_L load resistance, t_1 hold capacitor discharge time and t_2 pump capacitor discharge time to load (in parallel with the hold capacitor). The steady state average current can be calculated from the relation (5)

$$I_{ave} = \frac{Q_0 - Q(t_1)}{t_1} \quad (5)$$

where $Q(t_1) = C_h V_{out}(\infty)$ and Q_0 for the case in equation (3)

$$Q_0 = V_{DD} C_h C_p / (C_p + C_h (1 - \exp(-t_1/R_L C_h))) \quad (6)$$

A corresponding expression can be calculated for the case in eq. (4), too. Equation for the voltage ripple can also be derived. Equations (3) and (4) show that the output voltage has a steady state level between the negative of the supply voltage and ground. From the desired output current or load and the minimum allowable negative bias level one can calculate the requirements for the parameters of the charge

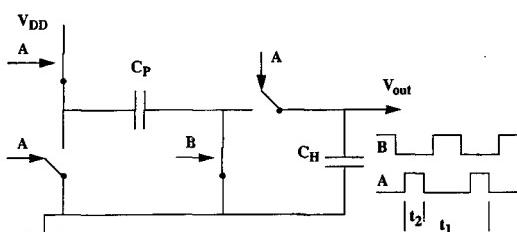


Fig. 1. The principle of a negative charge pump. The switches are shown in pulse A low, pulse B high position, C_p is the pump capacitor and C_h the hold capacitor.

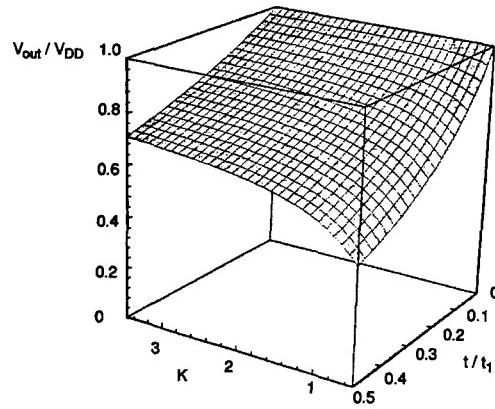


Fig. 2. The relative output voltage $V_{\text{out}}/V_{\text{DD}}$ as a function of the capacitance ratio $K = C_{\text{P}}/C_{\text{H}}$ and the time constant ratio t/t_1 , where $\tau = R_{\text{L}} C_{\text{H}}$.

pump. If we mark the load time constant $R_{\text{L}} C_{\text{H}} = \tau$ and the capacitance ratio $C_{\text{P}}/C_{\text{H}} = K$, we get for $\text{abs}(V_{\text{out}}) > 0.9V_{\text{DD}}$ the conditions $t_1 \ll \tau$ and $K > 1$ ($t_2 < t_1$ always). This sets the pumping frequency and the capacitance values. Figure 2 shows the relative output voltage as a function of the capacitance ratio K and the time constant ratio t/t_1 .

The preceding analysis is done for a resistive load. It can be applied also to switched capacitive loads on steady state

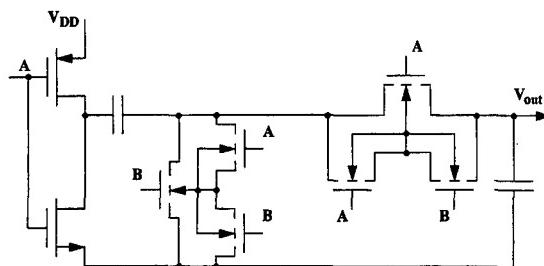


Fig. 3. The practical CMOS realization of the negative charge pump.

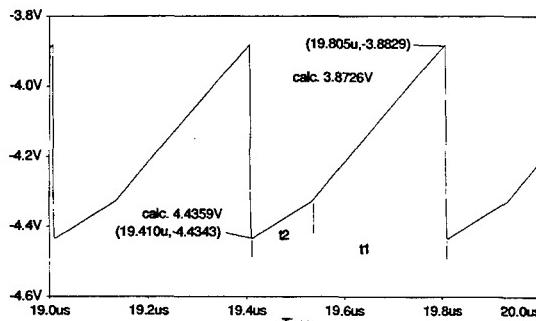


Fig. 4. The output voltage of an ideal charge pump with $C_{\text{H}} = C_{\text{P}} = 10\text{nF}$ and $R_{\text{load}} = 250\Omega$; and the calculated voltage values from eq. (4). The simulated average output current was 16.76 mA.

conditions with constant average current. This is the normal case with switched mode IGBT power drivers.

3. CMOS realization of a negative charge pump

The practical realization of the charge pump can be done with CMOS transistors, Figure 3. It requires either a n-substrate process or the possibility of floating NMOS transistors. Transistors T1 – T4 correspond the switches of Figure 1, and the function of the transistors T5–T8 is to keep the bodies of T3 and T4 in the most negative potential, to prevent latch-up. The on-resistance of the real switches sets another limit to the design. If the pump charging time (t_1-t_2), see Fig. 1, approaches the time constant $\tau_p = C_{\text{P}} R_{\text{on}}$, V_{DD} in eqs (2)–(6) will be under actual supply voltage. This sets an upper limit to the pump frequency. The substrates of the switching CMOS transistors must always be in the most negative potential to prevent latch-up. This determines the timing of t_1 and t_2 within a cycle, to give enough settle time for transistors T5–T8. Especially the edges of t_1 and t_2 must not be overlapping, because otherwise T3 and T4 and thereafter the hold capacitor C_{H} would be short-circuited.

The application of the negative charge pump described is limited to processes with n-substrate or suitable floating NMOS transistors. The design limitations come from the large area needed. For example a -5V 10mA source with 4.9V upper limit with $50\text{nF}/10\text{nF}$ C_{P} and C_{H} capacitors needs min. 3.3MHz pumping frequency. This set the maximum resistance of the main switches to 1.3Ω . Typical gate width would then be $15\text{--}40\text{mm}$ and area $0.6\text{--}1.4\text{mm}^2$. The fact that the outputs to the external capacitors cannot have protection components may cause reliability problems.

Figure 4 shows a simulated output voltage of the ideal negative charge pump in Fig. 1. The simulation has been done with PSPICE circuit simulation program. The switches have been ideal switches with $R_{\text{on}} = 1\Omega$ C_{H} and C_{P} were both 10nF and $R_{\text{load}} = 250\Omega$. The supply voltage V_{DD} was 5V and the pumping frequency was 2.5MHz . The nominal output current is $I_{\text{load}} = 5\text{V}/250\Omega = 20\text{mA}$. The simulation time has been long enough to reach the steady state conditions. The simulated upper and lower limit values are very close to the calculated ones.

References

1. di Cataldo, G., Palumbo, G., IEE Proceedings-G, **140** No. 1, 33 (1993).
2. Pataikuni, R., Carr, W. N., "Design and analysis of an optical-powered on-chip power supply." Proc. of the 32nd Midwest Symposium on Circuits and Systems (1989), p. 681.

Processing and Characterisation of an Etched Groove Permeable Base Transistor on 6H-SiC

Christer Fröjdः,* Göran Thungström,* Sofia Hatzikonstantinidou, Hans-Erik Nilsson* and C. Sture Petersson

Royal Institute of Technology, Department of Solid State Electronics, Electrum 229 S-164 40 Kista, Sweden

Received June 6, 1994; accepted in revised form July 5, 1994

Abstract

The Permeable Base Transistor (PBT) is generally considered as an interesting device for high speed applications. PBTs have been fabricated on Silicon and Gallium Arsenide by a number of groups. In this paper we reported on the fabrication of an etched groove PBT structure on 6H-SiC using Ti as contact metal for all electrodes. The devices have been characterised by DC-measurements. The transistors show the normal IV-characteristics for a such a device except for a parasitic series diode at the drain electrode. The breakdown voltage of the gate-drain diode is generally as high as around 60 V even without passivation of the sidewalls of the grooves.

1. Introduction

Silicon Carbide is an interesting material for high temperature, high frequency and high power devices. A summary of the properties of this material can be found in [1, 2]. Many devices may have been manufactured in SiC during the past years including diodes, MOSFETs and MESFETs. There is however not reported any working PBT device on SiC.

The PBTs are considered to perform at very high frequencies. Such a device on GaAs achieved a cut-off frequency of 200 GHz [3]. For PBTs on Silicon cut-off frequencies of 120 GHz have been simulated [4]. The measured cut-off frequencies for Silicon devices are however in the range of 25 GHz [5, 6].

PBTs could either be manufactured by etching of grooves for deposition of the gate metal or by overgrowing the gate grid with Silicon. The etched groove PBT is estimated to have better high frequency properties than the overgrown [7].

There is some confusion in the literature concerning the names of the electrodes of the PBT. However since the device essentially operates as a vertical short channel MESFET we have used the names source, gate and drain in this paper.

SiC offers a number of advantages for manufacturing of PBTs compared to silicon. The high saturation velocity for electrons in SiC predicts higher operating frequency than in silicon. Schottky contacts formed on SiC generally have much higher barriers than Schottky contacts to silicon. This will reduce the gate leakage current and in combination with the high breakdown field in SiC allow fabrication of PBTs with high channel doping. This is important since the maximum operating frequency of a PBT increases with its

transconductance and thus with the doping concentration in the channel.

We have used the analytical expression for the unity current gain frequency, f_T , for an over-grown PBT proposed in [8] to estimate the SiC PBT high frequency properties. The original model has been modified to be valid for an etched groove PBT and the final expression is:

$$f_T = \frac{g_{m, \text{sat}}}{2\pi C_{\text{tot}}} = \frac{v_{\text{sat}}}{2\pi L_{\text{eff}}}$$

$$L_{\text{eff}} \approx \frac{a}{\sqrt{\eta}} + \sqrt{\frac{2\varepsilon_0 \varepsilon_r (V_{bi} - V_{gs})}{qN_D} (2\alpha_s + \alpha_D \chi)}$$

$$\chi = \sqrt{1 + \gamma} \left(\frac{1}{1 + \gamma} + 1 \right)$$

$$\gamma = \frac{V_{ds}}{V_{bi} - V_{gs}}$$

where

f_T	= unity gain frequency
$g_{m, \text{sat}}$	= transconductance at saturation
C_{tot}	= total capacitance at the gate
v_{sat}	= saturation velocity for carriers
L_{eff}	= effective channel length, explained in text below
a	= line width of the grid, the grid spacing is $2a$
V_{bi}	= built in voltage of the gate diode
V_{gs}	= gate to source voltage
V_{ds}	= drain to source voltage
N_D	= doping concentration in the channel

The first equation in this set is the standard equation for f_T of a MESFET. The channel length of the ideal MESFET has, in this case, to be replaced by the effective channel length L_{eff} taking into account the two dimensional effects and the capacitance under the gate. The parameters η , α_D , α_s reflect the geometry of the grid and can be regarded as independent of material and doping levels. The values of these parameters are always in the range 0.5–1. Estimated f_T for a SiC PBT with grid line width $a = 0.2 \mu\text{m}$, channel doping $N_D = 1.5 \times 10^{18} \text{ cm}^{-3}$, $\alpha_D = 0.5$, $\alpha_s = 0.5$, $\eta = 0.5$, $v_{\text{sat}} = 2 \times 10^7 \text{ cm/s}$, $V_{ds} = 4.0 \text{ V}$ and $V_{gs} = 0.0 \text{ V}$ is 90 GHz. The same estimation for a Si PBT with the same parameters except $N_D = 1.0 \times 10^{17} \text{ cm}^{-3}$ and $v_{\text{sat}} = 1.0 \times 10^7 \text{ cm/s}$ yields $f_T = 32 \text{ GHz}$.

The high channel doping for the PBT on SiC has been chosen according to our results for schottky diodes on SiC in another study [9].

* Also: Mid-Sweden University, Dept. of Electronics, S-851 71 Sundsvall, Sweden.

A large part of the effective channel length is due to the capacitance under the gate fingers. If these capacitance could be removed then the estimated f_T increases to 550 GHz for the SiC PBT and 80 GHz for the Si PBT. The estimation is based on the drift diffusion model and no hot electron effects have been included. The transconductance is proportional to the square root of the doping level which gives a 3.87 times higher transconductance for $N_D = 1.5 \times 10^{18} \text{ cm}^{-3}$ compared to $N_D = 1.0 \times 10^{17} \text{ cm}^{-3}$. If the two times higher saturation velocity in SiC compared to Si is also included then the transconductance of the SiC device rises to 7.74 times the transconductance for a Si device with the same geometrical structure.

In this project we fabricated an etched groove PBT. The drain consists of finger shaped structures formed by dry etching of grooves. The width of the grooves is in the range 1–2 μm . The gate area is extended in order to allow characterisation by probing. A top view of a transistor is shown in Fig. 1. Transistors with four different sizes of the active area exist on the mask. The areas are 625, 2500, 5625 and $10000 \mu\text{m}^2$. Gate and source contacts are formed by self aligned metallisation. Ti is evaporated on to the wafer. The sidewalls of the grooves are then cleaned by wet etching in a HF solution. The method of self aligned silicidation used for PBTs on silicon [10] is more difficult to use in SiC technology since Schottky contacts formed by silicidation on SiC are generally less ideal than Schottky contacts formed by deposition of metal [11, 12].

An optimal wafer for this process is a n-type 6H-SiC bulk wafer with two n-type epitaxial layers on top of it. The first epitaxial layer, with low doping, is used for the channel of the transistor. The second layer, with high doping is used for source or drain contact depending on the polarity of the voltage in the channel. Unfortunately no such wafer was available at the time of processing so we had to use a wafer with a single epitaxial layer. The disadvantage of that is that the top contact will be rectifying. That contact then has to be used as drain causing high electric fields at the sidewalls. The structure of the transistor is shown in Fig. 2.

An advantage of this process is that the entire structure can be created using epitaxial layers. Thus doping by ion

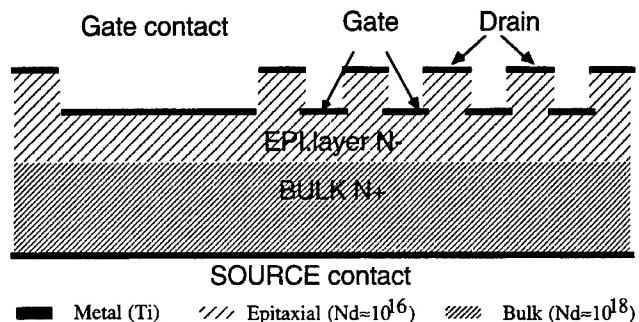


Fig. 2. Side view of a transistor showing the different layers. Since the wafer used lacked the upper epitaxial layer the top contact became rectifying. Thus the back contact had to be used as source and the top contact had to be used as drain, i.e. forward biasing the diode.

implantation, a technique which still has severe problems, could be avoided.

2. Processing

The starting material for this process was a n-type 6H-SiC wafer ($N_D = 3 \times 10^{18} \text{ cm}^{-3}$) with a n-type epitaxial layer on top of it. The wafer was supplied by Cree Research Inc. The doping of the epitaxial layer was specified to be $N_D = 3.6 \times 10^{15} \text{ cm}^{-3}$ and the thickness of the layer was specified to be 3 μm . The etch mask for formation of grooves was defined using standard photolithography. Titanium was evaporated on the wafer with a thickness of 200 \AA , then a 2000 \AA layer of Nickel was evaporated on top of the Ti layer. The metal mask pattern was then defined by lift-off. The Ti layer causes good adhesion to the SiC and protects the photoresist during evaporation of Ni. The Ni layer protects the Ti layer during etching.

Etching was done by RIE in a gas mixture of 50 sccm CF_4 , 10 sccm Ar and 5 sccm O_2 at a RF-power of 100 W during 1 h. The pressure in the chamber during etching was 15 mTorr.

The result of the etching was checked with SEM and the etch depth was measured to be 1 μm . (Fig. 3) The sidewalls were not exactly vertical. The deviation from the vertical axis was about 23°. This effect causes the finger widths to be larger than expected from the lithography.

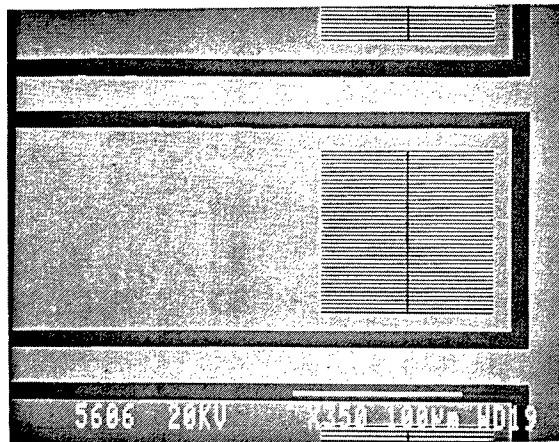


Fig. 1. SEM-picture showing the mask layout for one transistor. The dark areas are source (drain) and the light areas are gate. The gate is extended to the left in order to give space for probing. The source fingers are connected by the vertical bar in order to ensure that all fingers are contacted when probing.

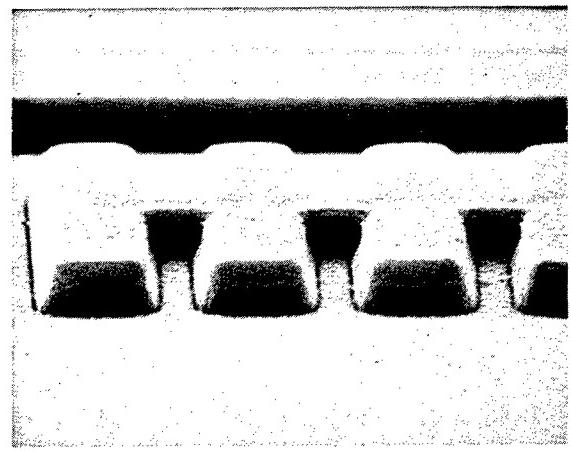


Fig. 3. SEM micrograph showing some transistor fingers after etching and metallisation. The nominal width of these fingers was 2 μm . Due to widening of the pattern during lithography and the non-vertical slopes from etching the finger width at the bottom of the groove is 3 μm .

After etching the structures were cleaned in an oxygen plasma for 10 min in order to remove any residues from the etching. Then 3000 Å of Ti was evaporated on top of the wafer using an electron beam evaporator. The base pressure before evaporation was about 10^{-8} Torr.

The evaporated metal film is much thinner on the sidewalls than on the top and bottom of the grooves. In order to clean the sidewalls, and thus isolate the gate from the drain, the samples were etched for 4 min in a 0.1% HF solution. The sheet resistance of the Ti layer had then increased by a factor of two indicating that the thickness of the Ti layer had decreased to half of its initial value. Electrical measurements confirmed that short circuits between gate and drain no longer existed.

Back side metallisation was done by evaporation of 3000 Å of Ti using the same equipment as for front side metallisation. Finally the samples were annealed at 515°C for 1 h. No passivation of the sidewalls was done.

3. Characterisation

After processing the transistors were characterised electrically by DC measurements. First the gate diodes were characterised by measuring *IV* and *CV* characteristics. Then *IV* characteristics for a number of transistors were measured and the results were compared to check the scaling properties.

CV-measurements were performed using a HP4279A *CV*-meter at a frequency of 1 MHz. A probe tip was placed on the metal area on the left side of the gate diode (Fig. 1). The back side of the wafer was used as the second electrode.

For reference the same measurements were done on a gate diode and on another diode with the same area but on an unetched part of the wafer. The results from the *CV*-measurements are presented in Fig. 4.

The border between the epitaxial layer and the bulk of the wafer can clearly be seen in the figure by looking at the different slopes of the curves. The measurements gave a doping concentration in the epitaxial layer of $N_D = 7 \times 10^{15} \text{ cm}^{-3}$, a Schottky barrier height of $\phi_B = 0.95 \text{ eV}$, an initial thickness of the epitaxial layer of $1.6 \mu\text{m}$ and an etching depth of $1 \mu\text{m}$, using a diode area of $26400 \mu\text{m}^2$.

IV-characteristics of the gate diodes were measured in a computerised measurement system using HP3245A Univer-

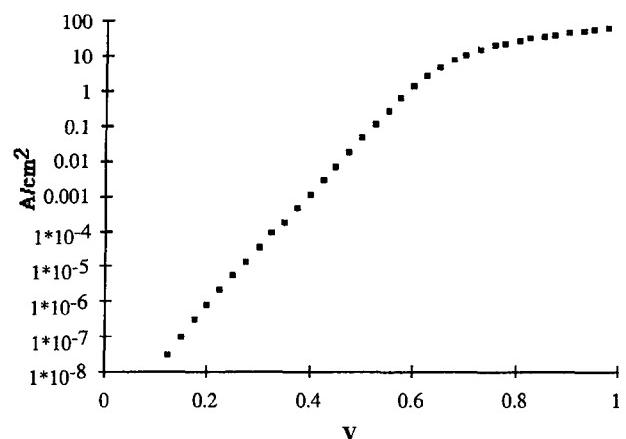


Fig. 5. Forward current density as a function of forward voltage for a gate diode. The ideality factor is 1.05.

sal Source and HP3458A Multimeter. The results of these measurements are presented in Fig. 5. From these results a Schottky barrier height of $\Phi_B = 0.96 \text{ eV}$ and an ideality factor of 1.05 were extracted. The diodes show good rectifying behaviour and the reverse leakage currents were always below $1 \mu\text{A}$ at reverse voltages of up to 30 V.

The same measurement set-up was also used for measuring the *IV*-characteristics of the transistors. The top electrode contact was used as drain and the back side of the wafer was used as source. During the measurements the source was grounded and the voltages of the gate and the drain were varied. The results for a typical transistor are presented in Fig. 6. The drain current is measured for a number of different gate voltages. The gate voltage varies in the range 0–20 V and the drain voltage varied in the range 0–10 V. The diode at the drain contact causes the drain current to be almost zero for drain voltages below 1.5 V.

The dependence of drain current on active area was checked by measuring the *IV*-characteristics for transistors with the same grid dimensions but with different active areas. Drain current scaled well with active area for all four sizes of transistors.

The dependence of drain current on grid dimensions was checked by measuring *IV*-characteristics of transistors with

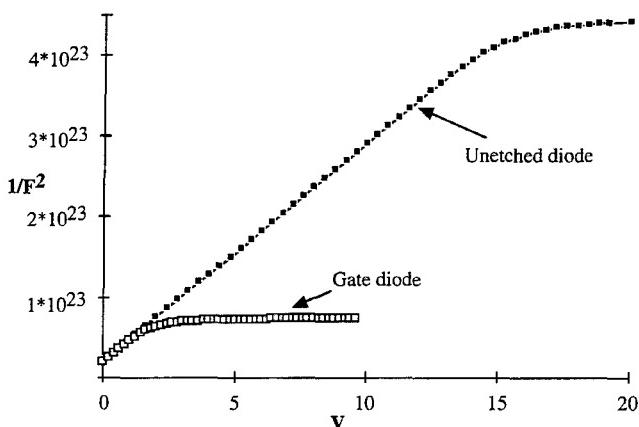


Fig. 4. $1/C^2$ as a function of reverse bias for a gate diode and an unetched diode with the same area. The border between the epitaxial layer and the bulk is at the knee in the curves. Etching depth and thickness of the epitaxial layer can be calculated from these curves.

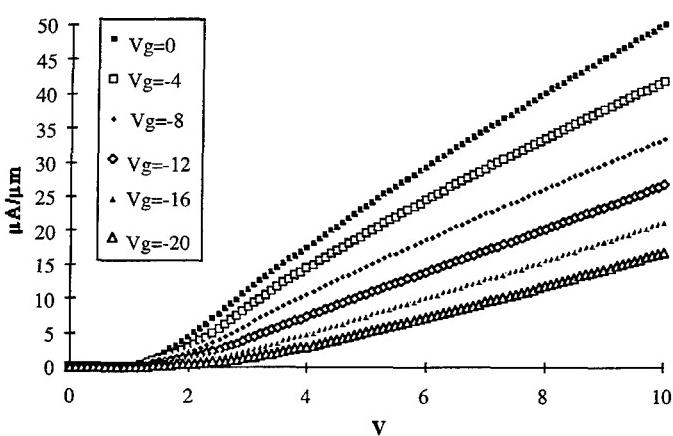


Fig. 6. *I/V* characteristics for a typical transistor. The effective grid width is $2.5 \mu\text{m}$. The drain current is expressed as current (in μA) per μm finger length.

different grid widths. In this case the scaling was not as exact as for transistors with different areas. This can probably be explained by uncertainties in the lithography and effects of the non-vertical sidewalls.

In this process no passivation of the sidewalls after etching of the grooves was done. This was expected to cause some problems with high leakage currents and diode breakdown especially since the top electrode had to be used as drain leading to high electrical field at the sidewalls. However most of the gate-drain diodes could withstand at least 60 V.

4. Simulations and discussion

In order to compare our results with theoretical results the structure was also simulated with MEDICI. The values used in the simulation were $N_D = 7 \times 10^{15} \text{ cm}^{-3}$, epi layer thickness $1.6 \mu\text{m}$ and finger width $2.5 \mu\text{m}$. Material dependent parameters were essentially extracted from [13].

The resulting *IV*-characteristics from the simulation are presented in Fig. 7. Scaling is done to give the drain current in microamperes per micrometer channel length. The drain current for the simulated transistor is higher than the measured drain current. This can be explained by series resistance at the source and drain contacts and the diode behaviour of the drain contact.

The PBT is one of the most simple three terminal devices to fabricate on 6H-SiC. With proper starting material only dry etching and metallisation has to be used. It is also very simple to fabricate transistors with different threshold voltages in the same process since the threshold voltage, for

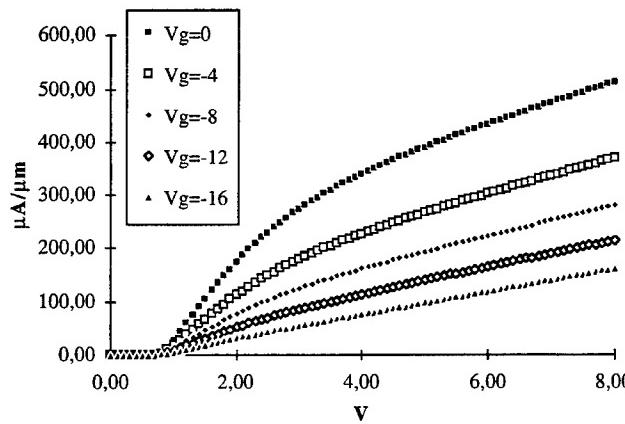


Fig. 7. Simulated *IV*-characteristics for the same transistor as in Fig. 6.

transistors with the same doping, is only controlled by the finger width.

5. Conclusions

We have fabricated PBTs on 6H-SiC using Ti as contact metal. The transistors show the expected behaviour. The transistors are able to handle high voltages without any sidewall passivation. This is very promising for further development of SiC permeable base transistors.

The material properties of SiC makes it suitable for obtaining a PBT with high f_T . We have also theoretically showed that cut off frequencies in the range of 500 GHz could be achieved with the proper material and an optimised process.

Acknowledgements

We are very grateful to Janos André at KTH-HLB for doing the etching and for valuable advice concerning the lithography.

References

1. Ivanov, P. A. and Chelnokov, V. E., *Semicond. Sci. Technol.* **7**, 863 (1992).
2. Davis, R. F., Kelner, G., Shur, M., Palmour, J. W. and Edmond, J. A., *Proceedings of IEEE*, **79**, 677 (1991).
3. Gruhle, A. and Benking, H., *IEEE El. Dev. Lett.*, **11**, 165 (1990).
4. Oshima, T., Nakamura, N., Nakagawa, K., Yamagushi, K. and Miyao, M., *Proc. IEDM*, **33** (1991).
5. Rathman, D. D. *et al.*, *Silicon Permeable Base Transistors*, 16th Int. Conf. Solid St. Dev. and Mat. 1984. Ext. Abstracts pp. 305–308.
6. van Rijs, F., Oostra, D. J., Rooij-Mulder, J. M. L. and Timmering, C. E., *Permeable Base Transistors with Schottky and Junction Gates*. ESSDERC 1993.
7. Vojak, B. A. *et al.*, *IEEE Trans. El. Dev.*, **30** 877 (1983).
8. Nilsson, H. E., Sannemo, U. and Pettersson, C. S., *Nordic Semiconductor Meeting 1994*, *Physica Scripta T54*, 159 (1994).
9. Fröjd, C., Thungström, G., Nilsson, H. E. and Pettersson, C. S., *Nordic Semiconductor Meeting 1994*, *Physica Scripta T54*, 169 (1994).
10. Hatzikontanidou, S., Nilsson, H. E., Fröjd, C. and Pettersson, C. S., *Nordic Semiconductor Meeting 1994*, *Physica Scripta T54*, 226 (1994).
11. Lundberg, N., Zetterling, C.-M. and Östling, M., *Appl. Surf. Sci.* **73** 316 (1993).
12. Bathnagar, M. *et al.*, *IEDM92*, pp. 789–792, 1992.
13. Helbig, R., *5th Int. Symp. Power Semic and IC's*, pp. 6–11 (1993).

The Influence of Emitter Properties on the Heat Generation in SiC and Si PIN Diodes Under Forward Conduction

Olof Tornblad,¹ Bo Breitholtz,^{1,2} Mikael Östling^{1,3} and Ulf Lindefelt^{2,4}

¹ Royal Institute of Technology, Department of Electronics, Solid State Electronics, Electrum 229, S-164 40 Kista-Stockholm, Sweden

² ABB Corporate Research AB, S-721 78 Västerås, Sweden

³ Present Address: Stanford University, CIS, Stanford, CA 94305, USA

⁴ Material Physics, Dept. of Physics and Measurement Technology, University of Linköping, S-581 83 Linköping, Sweden

Received June 6, 1994; accepted June 14, 1994

Abstract

This study focuses on the impact of the emitters on the forward voltage drop (which affects the heat generation) in Si and 6H-SiC PIN 5 kV diodes. The influence of the Shockley-Read-Hall (SRH) lifetime on the forward voltage drop was investigated. It was found that the forward voltage drop of the 5 kV Si diode depends much stronger on the SRH lifetime than the 5 kV SiC diode. The influence of minority carrier transport in the highly doped emitters on the forward voltage drop in 5 kV Si and SiC PIN diodes was quantified. Additionally, two-dimensional simulations of a mesa-etched 5 kV PIN diode in 6H-SiC show a large impact of anisotropy on the forward voltage drop at high current densities. These simulations indicate that large advantages with respect to forward voltage drop could be achieved in mesaetched 5 kV SiC PIN diodes by choosing the substrate in a direction parallel to the c-axis. This result is interesting since most substrates commercially available today have the polished surface perpendicular to the c-axis.

1. Introduction

Heat generation in power devices is responsible for a substantial amount of the cost for power electronic systems. Therefore it is of great interest to the power industry to try to reduce the dissipation of heat in power devices under static and dynamic conditions. The PIN diode represents the fundamental structure for several power devices under forward conduction. Recombination at high current densities will predominate in the highly doped n⁺ and p⁺ emitters and will influence the injection level of the electron-hole plasma in the n-base, which affects the forward voltage drop in a Si PIN-diode [1–4]. The heat generation in a Si diode will thus strongly depend on the properties of the emitters. The progress of numerical simulation of semiconductor devices during the last years has made it possible to study much more in detail the influence of a variety of physical models on device performance.

In the field of power electronics silicon carbide (SiC) is attracting a lot of attention. This is because of its high critical electric field, large bandgap and high thermal conductivity. A lot of the present research in the area focuses on growth and characterization of this material. However, it is of great interest to explore the characteristics that can be expected for future SiC devices. From general considerations, equations for anisotropic properties of semiconductors such as SiC have been derived [5]. These equations have also been implemented in a simulation code [6], which enables us to investigate the impact to anisotropy in SiC power devices. Experimental work [7–9] has made it possible to calibrate some of the physical models to measurements.

The present study focuses on the impact of the emitters on the forward voltage drop (which affects the heat generation) in Si and 6H-SiC PIN 5 kV diodes. Figure 1 shows the structure, which has been used in the one-dimensional simulations. The emitter doping profiles were chosen as nearly abrupt. W_m denotes the width of the lowly doped middle region, W_{en} and W_{ep} the depth of the n⁺ and p⁺ emitters respectively. The emitter doping concentrations are labelled N_{en} (n-side) and N_{ep} (p-side). Finally, a demonstration of the impact of anisotropy in mobilities, dielectric constant and ionization coefficients on the forward voltage drop of 6H-SiC bipolar devices is presented; 2D simulations of mesa-etched 5 kV PIN-diodes with two different crystal orientations have been chosen for this illustration. This structure is shown in Fig. 2.

2. Theory

2.1. Recombination in PIN diodes

The influence of the emitter properties on the forward voltage drop (and adherent heat generation) in a PIN-diode is partly caused by the injection of minority carriers into the emitters. A high injection of minority carriers into the emitters affects the injection level of the electron-hole plasma in the lowly doped middle region, which will influence the forward voltage drop. The minority carrier current injected into an n⁺ emitter can be written as,

$$J_p = J_{\text{bulk}} + J_{\text{surf}} = q \int_0^W U \, dx + q s_p (p_s - p_{s0}) \quad (1)$$

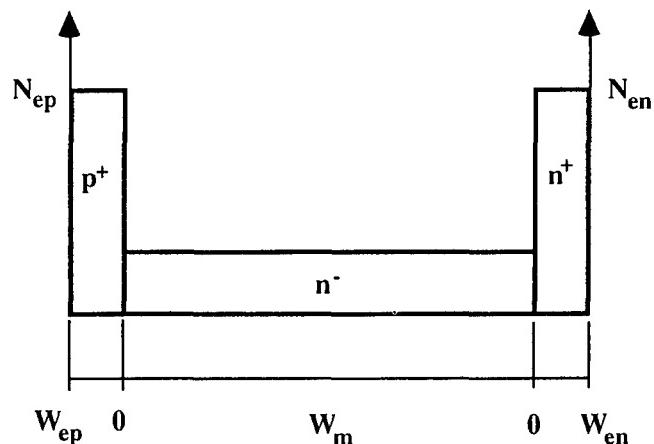


Fig. 1. PIN-diode in 1D.

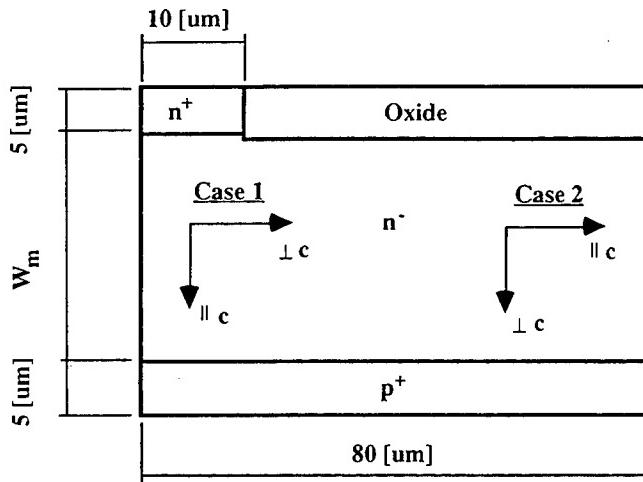


Fig. 2. Mesa-etched PIN-diode for demonstration of anisotropy,

where J_{bulk} and J_{surf} are the contributions to the minority current from carriers recombining in the emitter bulk and at the surface, i.e. at the emitter contact, respectively. W denotes the emitter depth (including the space-charge region), U the recombination rate of electron-hole pairs, s_p the surface recombination velocity at the emitter contact, p_s the hole concentration at the emitter contact and p_{s0} the hole concentration at the emitter contact in thermal equilibrium.

Two types of emitters can be classified [10] with the help of the transparency factor,

$$\alpha_t = J_p(W)/J_p(0) \quad (2)$$

where W was defined to denote the location of the emitter contact and 0 the emitter edge of the space-charge layer in the quasi-neutral part of a diffused emitter. It demonstrates the fraction of carriers recombining at the emitter contact as opposed to those recombining in the emitter bulk. For opaque emitters, almost none of the minority carriers that are injected into the emitter will reach the emitter contact and $\alpha_t \approx 0$. By contrast, for transparent emitters nearly all minority carriers that are injected into the emitter reach the emitter contact and $\alpha_t \approx 1$. Because of this there is a large difference in the influence of the surface recombination velocity at the emitter contact for the two emitter types.

2.2. Physical models for simulation

In this study all simulations were done with the simulation program MEDICI [11], currently under development to account also for anisotropic properties in semiconductors.

In the present study, simulations have been done within the drift-diffusion approximation with a constant temperature (thus excluding the heat conduction equation). Consequently, the set of equations used is Poisson's equation and the continuity equations for electrons and holes. For 6H-SiC, the dielectric constant and low-field mobility are diagonal tensors of rank 2 with two elements equal, different from the third [5]. The following physical models were used for the Si simulations: Dorkel-Leturq mobility model [12], SRH recombination with doping dependent lifetimes [13], Auger recombination with constant Auger coefficients, (using the Dziewor-Schmid values $C_n = 2.8 \cdot 10^{-31}$, $C_p = 1.0 \cdot 10^{-31}$ [14]) band gap narrowing [15] and Fermi-Dirac statistics.

For the 6H-SiC simulations mobility was modelled according to [9]. At 300 K the anisotropy takes the values,

$$\mu_{n,\perp} = \mu_n \quad \mu_{p,\perp} = \mu_p \quad (3a)$$

$$\mu_{n,\parallel} = \mu_n/4.914 \quad \mu_{p,\parallel} = \mu_p \quad (3b)$$

where \parallel denotes a direction parallel to the c -axis and \perp a direction perpendicular to the c -axis. Incomplete ionization of donors and acceptors has been accounted for. Doping with nitrogen leads to a donor which has three different energy levels below the conduction band. The ionization energies are [17]

$$E_c - E_{d1} = 81 \text{ meV}, E_c - E_{d2} = 137.6 \text{ meV},$$

$$E_c - E_{d3} = 142.4 \text{ meV} \quad (4)$$

In actual simulations and if one is not concerned with the dynamic effects of incomplete ionization, it can be demonstrated that these three energy levels to a very good approximation can be lumped together and replaced by a single level at

$$E_c - E_d = 100 \text{ meV} \quad (g_d = 2) \quad (5)$$

Doping with Al leads to one acceptor level at [18]

$$E_a - E_v = 200 \text{ meV} \quad (g_a = 4) \quad (6)$$

No Auger coefficients are available for any polytype of SiC. A very crude order of magnitude choice would be to use the Dziewor-Schmid values for Si ($C_n = 2.8 \cdot 10^{-31}$, $C_p = 1.0 \cdot 10^{-31}$, [14]). Impact ionization coefficients were modelled as

$$\alpha_{n,p} = \alpha_{n,p} \exp(-b_{n,p}/E_j) \quad (7)$$

where E_j is the projection of the electric field in the direction of current. From references [7] and [8] we have

$$\alpha_{n,\parallel} = \alpha_n \quad \alpha_{p,\parallel} = \alpha_p \quad (8a)$$

$$\alpha_{n,\perp} = \alpha_p/3.5 \quad \alpha_{p,\perp} = \alpha_p \quad (8b)$$

At 300 K these parameters take the values (as can be deduced from [7]):

$$a_n = 4.95 \cdot 10^6 \text{ cm}^{-1} \quad a_p = 2.16 \cdot 10^7 \text{ cm}^{-1} \quad (9a)$$

$$b_n = 2.58 \cdot 10^7 \text{ V/cm} \quad b_p = 1.90 \cdot 10^7 \text{ V/cm} \quad (9b)$$

The components of the low frequency dielectric constant have the values [19] $\epsilon_\perp = 9.66$ and $\epsilon_\parallel = 10.03$. Finally, the bandgap for 6H-SiC at $T = 300$ K is 2.95 eV [20], whereas the effective masses have been set equal to the free electron mass. This is generally of minor importance, but reflects the fact that published values of the effective masses differ substantially.

3. Method of comparing PIN-diodes in Si and SiC

Comparison of the heat generated in PIN-diodes in Si and SiC under forward conduction has been made. In order to compute the total heat generated by these diodes for a specific application it would be necessary to take into account both static and dynamic losses for that specific circuit and driving conditions. Here, however, only the differences during forward conduction are demonstrated. The criteria for comparing the two PIN-diodes in Si and SiC was to design them for the same reverse blocking capability. Non punch-through designs for 5 kV were chosen.

4. Results and discussion

Simulations of the structure in Fig. 1 in the reverse direction with a breakdown voltage of 5 kV yielded a doping concentration in the middle layer of $N_d = 2 \cdot 10^{13} \text{ cm}^{-3}$ for the Si case; this gives a width of $W_m = 572 \mu\text{m}$ of the lowly doped middle region for a non punch-through design. The corresponding values for the SiC diode were found to be $N_d = 3.5 \cdot 10^{15} \text{ cm}^{-3}$ with $W_m = 40 \mu\text{m}$ (electric field parallel to c -axis) and $N_d = 2.6 \cdot 10^{15} \text{ cm}^{-3}$ with $W_m = 45 \mu\text{m}$ (electric field perpendicular to c -axis). Figure 3 shows how the forward voltage drop varies with high injection SRH lifetime ($\tau_n + \tau_p$), obtained by numerical simulation for the Si diode and the 6H-SiC diode (with electric field along the c -axis) for two different current densities. Emitter doping concentrations $N_{en, ep} = 5 \cdot 10^{19} \text{ cm}^{-3}$ and emitter depths $W_{en, ep} = 10 \mu\text{m}$ were chosen for this simulation. The two diodes show quite different behaviour. The forward voltage drop of the Si diode depends much stronger on SRH lifetime than the SiC diode. If we consider a current density of 100 A/cm^2 (a typical operating condition), we note that up to approximately $4 \mu\text{s}$ the SiC diode shows the lowest forward voltage drop. This is because of a much smaller voltage drop over the middle region in the SiC diode compared to Si. On the other hand, for lifetimes greater than approximately $4 \mu\text{s}$ the Si diode will have the lowest forward voltage drop. This seems reasonable since the total voltage drop over the emitters in the Si diode is smaller than in the SiC diode due to a smaller built-in potential of the Si emitter junctions (because of the bandgap difference). Thus, when the voltage drop over the middle region in the Si diode is small enough, the forward voltage of the Si diode will be the lowest. Increasing the current enhances the importance of the middle region. As can be seen in Fig. 3, for a current density of 1000 A/cm^2 the forward voltage drop of the SiC diode is lower than the of Si for lifetimes shorter than approximately $60 \mu\text{s}$. SRH lifetime will be a much less critical parameter for the forward voltage drop of the SiC diode than its Si counterpart; this means that it is possible to design a 5 kV SiC diode for high switching frequencies and still keep a reasonably low forward voltage drop. By contrast, for high lifetimes (i.e. low frequency applications) the Si diode offers the lowest forward voltage drop. In order to compare the total heat generated by the diodes for a certain frequency it is necessary to take into account the dynamic losses.

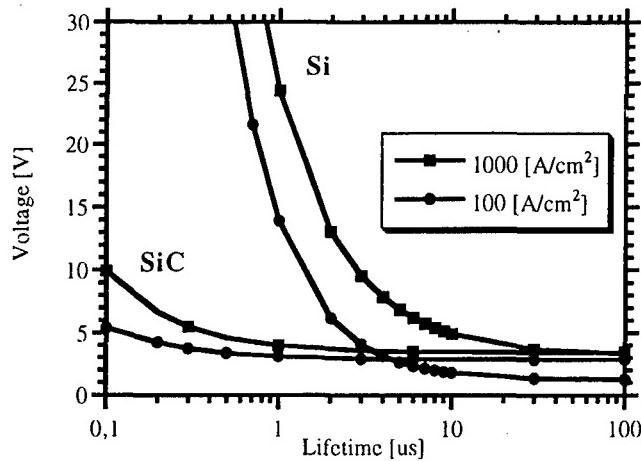


Fig. 3. Forward voltage drop versus high injection SRH lifetime ($\tau_n + \tau_p$).

Emitter doping concentration and depth of the emitters were varied in order to investigate what effect the minority carrier transport in the highly doped emitters has on the conductivity modulation and the resulting forward voltage drop of the diodes; the SRH lifetime was set to $4.4 \mu\text{s}$, which corresponds to a high frequency Si diode. With this lifetime the two diodes will have the same forward voltage drop at 100 A/cm^2 for the chosen emitters, and both are designed for 5 kV. Figure 4 shows the results of varying the n^+ emitter for the Si and SiC diodes at a current density of 100 A/cm^2 . The simulations were done for two surface recombination velocities for the minority carriers at the emitter contact; $s = \infty$ corresponds to a metal contact and $s = 10^3 \text{ cm/s}$ is a possible value for a poly-Si emitter [16]. The influence of the emitter parameters is much stronger for the Si diode. This agrees well with the conclusion above, that the voltage drop over the middle region in the SiC diode will be smaller than for the Si diode for a lifetime of $4.4 \mu\text{s}$. An emitter with $s_p = 10^3 \text{ cm/s}$ is always better or equally good as the emitter with $s_p = \infty$. For Si emitters with an infinite surface recombination velocity a high emitter doping and depth is required to minimize the minority carrier current. A Si emitter with $s_p = 10^3 \text{ cm/s}$ should be as thin as possible and lightly doped; Fig. 4(a) suggests a doping of about $5 \cdot 10^{17} \text{ cm}^{-3}$ for the studied Si diode. These principles agree well with results presented by del

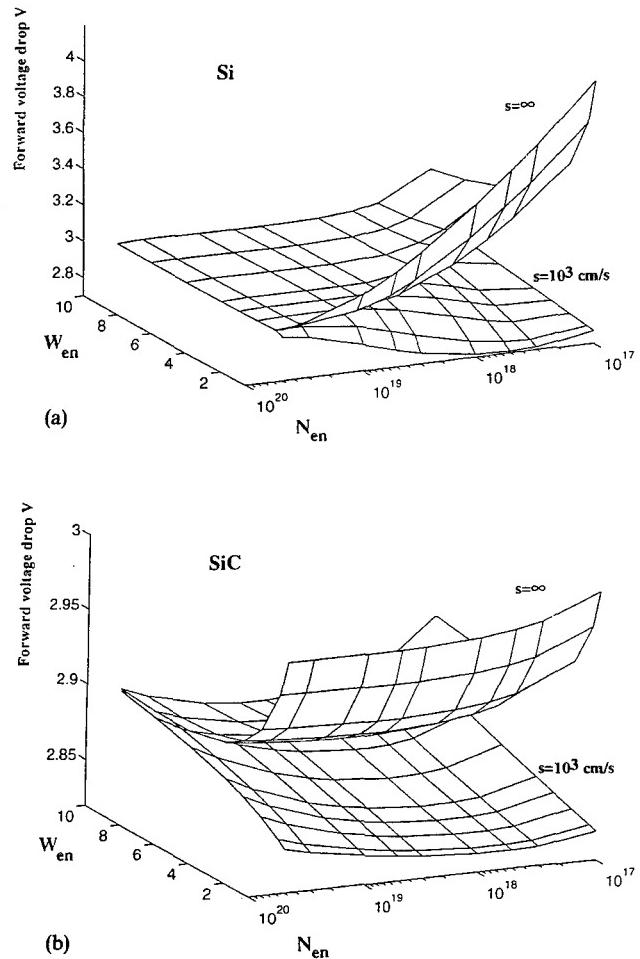


Fig. 4. Dependence of forward voltage drop on emitter doping concentration and emitter depth for the n^+ emitter. The lower curve planes are for $s = 10^3 \text{ cm/s}$ and the upper for $s = \infty \text{ cm/s}$; (a) 5 kV Si PIN diode, (b) 5 kV SiC diode. Note the different scaling of the z-axes.

Alamo [10] for diffused emitters. For the SiC emitter [Fig. 4(b)] the same dependencies are observed although much weaker. This tells us that the influence of the emitter doping, emitter depth and surface recombination velocity on the conductivity modulation in the investigated 5 kV 6H-SiC diode is more or less negligible with respect to forward voltage drop.

Figure 5 shows the results of 2D simulations of a mesa-etched 6H-SiC diode (defined in Fig. 2) for two *c*-axis orientations. Both diodes were designed for 5 kV through 1D simulations in the reverse direction. In Figs 5(a) and (5b) we note the large difference in the shape of the electron-hole plasma for the two crystal orientations. For case 1 (see Fig. 2), a lot of carriers are injected along the right part of the p^+

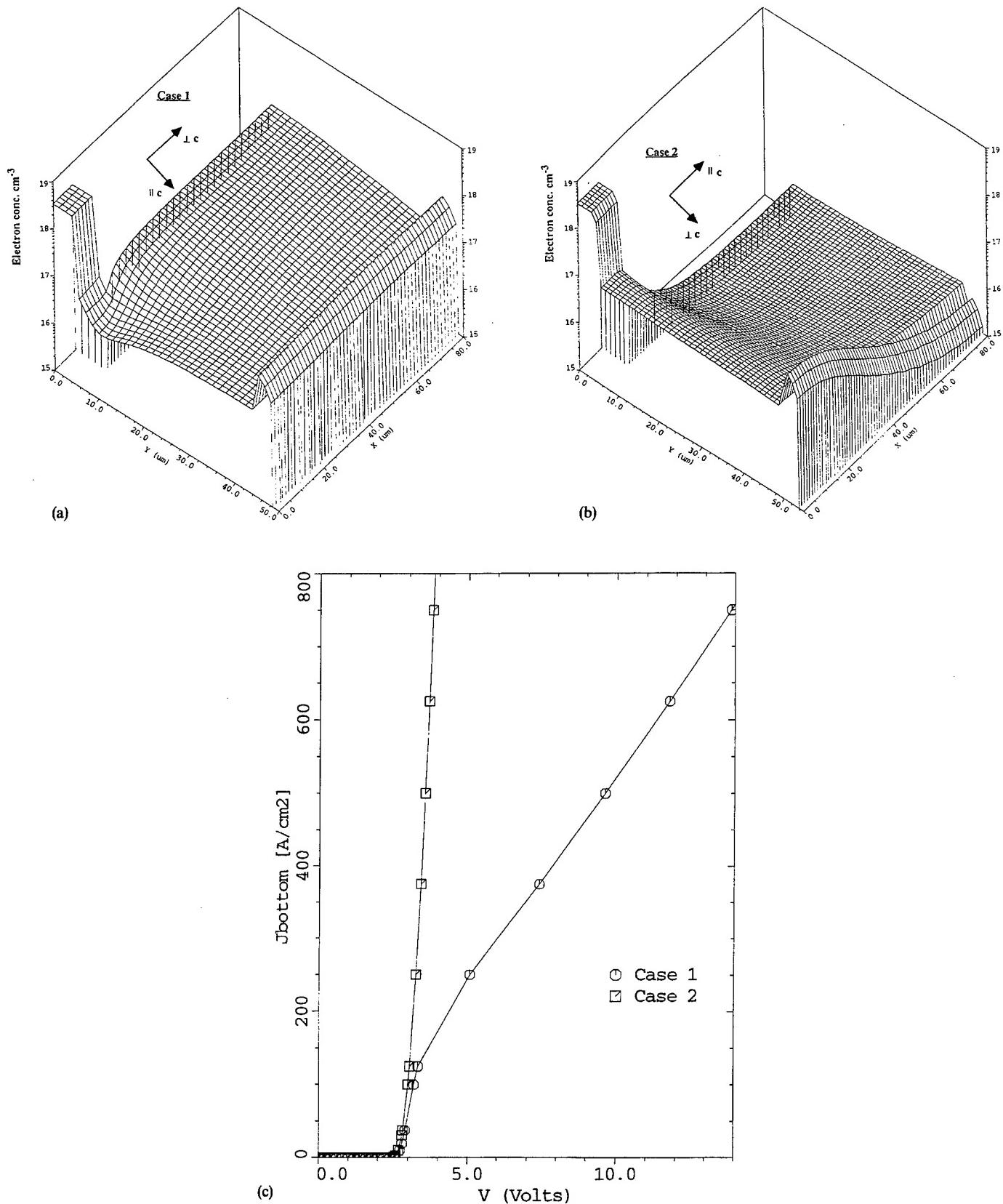


Fig. 5. Two 5kV SiC PIN diodes with different *c*-axis orientation; (a) Case 1 in Fig. 2. (b) Case 2 in Fig. 2. (c) IV-characteristics for the different orientations.

emitter contact. For case 2, almost all injected carriers are confined to the left part of the middle region. In Fig. 5(c) the resulting IV-characteristics are shown, suggesting that case 2 would be advantageous for low heat generation during forward conduction. This result is interesting for the fabrication of vertical bipolar power devices, since most 6H-SiC substrates available today are delivered with the *c*-axis perpendicular to the polished surface.

Conclusions

The impact of minority carrier transport in the highly doped emitters on the forward voltage drop (i.e. heat generation) in 5kV PIN diodes in Si and SiC under forward conduction was quantified. The importance of conductivity modulation is much less pronounced in a 5kV SiC PIN diode compared to the Si case. The forward voltage drop in the 5kV SiC diode is therefore less dependent on lifetime and also on current than its Si counterpart. Consequently, 5kV PIN-diodes in SiC will be a good choice for high currents and switching frequencies, whereas 5kV PIN-diodes in Si still offer the lowest forward voltage drop for low frequency applications.

2D simulations of a mesa-etched 5kV PIN diode in 6H-SiC show a large impact of anisotropy on the forward voltage drop at high current densities. These simulations indicate that large advantages with respect to forward voltage drop could be achieved in mesa-etched 5kV SiC PIN diodes by choosing the substrate in a direction parallel to the *c*-axis. This result is interesting since most substrates commercially available today have the polished surface perpendicular to the *c*-axis.

References

- Herlet, A., Solid St. Electron., **11**, 717 (1968).
- Adler, M. S., IEEE Trans. Electr. Dev., **ED-25**, 16 (1978).
- Burtscher, J., Dannhäuser, F. and Krausse, J., Solid St. Electron. **18**, 35 (1975).
- Berz, F., Cooper, R. W. and Fagg, S., Solid St. Electron., **22**, 293 (1979).
- Lindefelt, U., accepted for publication in J. Appl. Phys. (manuscript # JR-9605).
- ABB-TMA-IMC joint development.
- Konstantinov, A. O., Sov. Phys. Semicond. **23**, 31 (1989).
- Dimitriev, A. P., Konstantinov, A. O., Litvin, D. P. and Sankin, V. I., Sov. Phys. Semicond. **17**, 686 (1983).
- Schaffer, W., Negley, G., Irvine, K. and Palmour, J., to be published in Mat. Res. Soc. Symp. Proc., **339** (1994).
- del Alamo, J., Ph.D. Dissertation, Stanford Electronics Laboratories, Stanford University, Stanford, Calif., Nov. 1985.
- MEDICI is a trademark of Technology Modeling Associates Inc.
- Dorkel, J. M. and Leturq, Ph., Solid St. Electron. **24**, 821, (1981).
- Roulston, D. J., Arora, N. D. and Chamberlain, S. G., IEEE Trans. Electr. Dev., **ED-29**, 284 (1982).
- Dziewior, J. and Schmid, W., Appl. Phys. Lett. **31**, 346 (1977).
- Slotboom, J. W., Solid St. Electron. **20**, 279 (1977).
- Roulston, D. J., "Bipolar Semiconductor Devices" (McGraw-Hill Series in Electrical Engineering, New York 1990), p. 200.
- Suttrop, W., Pensl, G. and Choyke, W. J., J. Appl. Phys. **72**, 3708 (1992).
- Suttrop, W., Pensl, G. and Lanig, P., Appl. Phys. **A51**, 231 (1990).
- "Semiconductors, Group IV Elements and III-V Compounds in Data in Science and Technology" (Edited by O. Madelung) (Springer-Verlag 1991).
- Edmund, J. H. et al., in Proceedings from "First International High Temperature Electronics Conference", Albuquerque, New Mexico, 500 (1991).

High-Voltage Silicon Carbide Rectifiers – Results of Experiments and Simulation

L. P. Ramberg, S. Savage, U. Gustafsson and A. Schöner

IMC Industrial Microelectronics Center, P.O. Box 1084, S-164 21 Kista, Sweden

Received May 24, 1994; accepted in revised form July 5, 1994

Abstract

Low reverse leakage silicon carbide pin rectifier diodes with a breakdown voltage reaching 1100 V are experimentally shown to have acceptably low forward voltage drops, dominated by the built-in voltage. Numerical simulations of the experimental structure, using a measured carrier lifetime of 25 ns, validate the existence of a conductive plasma in the lowly doped base layer, despite a poor injection efficiency, related to the incomplete ionization of the aluminium acceptor. Simulation also indicates the need for a lifetime of ~ 100 ns in a 3.0 kV device.

1. Introduction

Silicon carbide (SiC) has been identified as a near-ideal semiconductor for power devices, owing mainly to its dielectric strength, thermal conductivity and band-gap [1]. The dielectric strength is extraordinary: at 3 MV/cm, the critical field acceptable before avalanche breakdown is ten times higher than in silicon. This allows a layer that blocks a certain voltage to be ten times thinner than in silicon, lowering the forward voltage drop. The thermal conductivity is about three times that of silicon: at 5 W/Kcm, it is similar to that of copper. The band-gap, lastly, is wide, ranging from 2.4 to 3.2 eV depending on the sort, or polytype, of silicon carbide. In the case of 6H SiC, the polytype of SiC used in this study, the band-gap is 3.0 eV, in principle permitting temperatures above 1000 °C in typical power devices.

Additional merits of SiC for power devices are the chemical inertness and mechanical robustness, which may facilitate packaging, and the radiation hardness.

While majority carrier devices such as Schottky diodes can be employed at fairly high voltages, in theory as high as 5 kV at room temperature [2], pin-structures that rely on double injection and conductivity modulation are necessary for achieving low conduction losses in high-voltage device designs operating at higher current densities and at higher temperatures [3]. This poses requirements on minority carrier lifetimes, emitter efficiencies and other properties that are little explored in SiC. The aim of the present work is to investigate experimentally and by means of simulation the attainability of conductivity modulation in SiC pin diodes, and thus their feasibility for high-voltage rectification with low on-state losses.

2. Device fabrication

Vertical p⁺n⁻n⁺ diodes were fabricated using commercial CVD-grown epitaxial SiC of the 6H polytype. The material was purchased in 1" wafers from Cree Research Inc [4].

The layer structure is shown in Table I, where the dopings are measured by a capacitance-voltage technique by the supplier.

The device geometry, a straightforward mesa with a somewhat retracted p-type contact on top, is shown schematically in Fig. 1. The diameter of the mesas are 50, 100 and 200 μm measured at the top. The mesas are about 1.2 μm tall, i.e. the p⁺ layer is etched through with a nominal 2000 Å margin. Etching is done with RIE using a CF₄/Ar/H₂ plasma at a rate of 1.0 μm/h. SEM reveals a 20° slope of the sidewalls.

The p-type contact consists of a sandwich of AlTi plus TiW_xAu deposited after a 300 s anneal at 950 °C in an RTA system. The n-contact consists of a NiTiPtAu, deposited over the entire back side. Before the deposition of the bottommost contact layer, a short in-situ Ar sputter is used to clean the SiC surface.

3. Results

p-type contact resistivity was measured with the transmission line method (TLM) to reach a low 2×10^{-5} ohm cm². n-type contact resistivity was not measured since the large size of this contact makes resistivity non-critical to our devices.

Table I. Wafer layer structure

Purpose	Thickness (μm)	Doping (cm ⁻³)
p ⁺ contact layer	1	1.1E18
n ⁻ base layer	9	7.9E15
n ⁺ buffer layer	0.5	>1E19
substrate	300	2E18

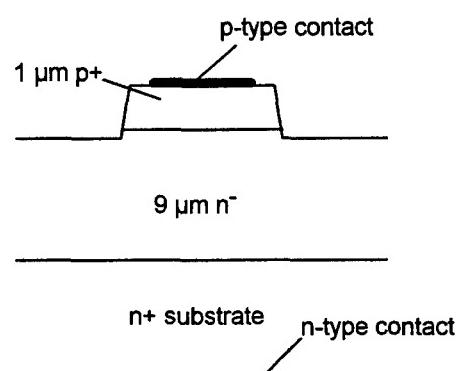


Fig. 1. Cross-sectional schematic view of a device.

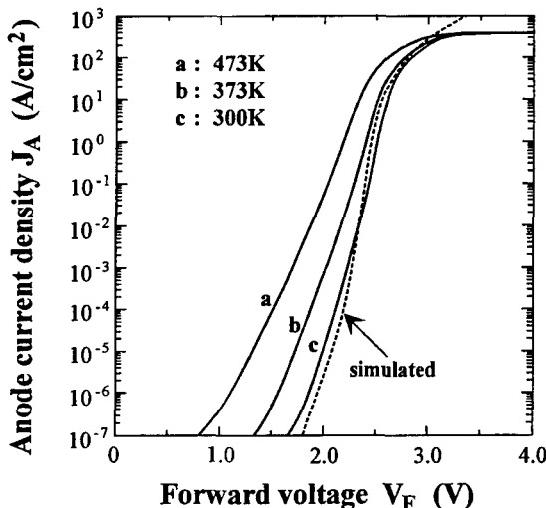


Fig. 2. Current density vs. forward voltage measured at three different temperatures, and the simulated characteristic for room temperature.

The highest breakdown voltage observed was $V_{br} = 1100$ V, with a full 60% of the 200 μm diameter devices being able to sustain 1000 V. While this is lower than the highest value reported [5], $V_{br} = 2200$ V, it is a high fraction of the calculated bulk breakdown voltage in the absence of edge effects, $V_{br} = 1700$ V. Measured in an SF₆ gas ambient, breakdown luminescence is clearly observed from bright spots around the periphery of the device, indicating current filamentation. About 1 mA of avalanche current could be passed through a 200 μm device, before destructive flashovers occur. Flashovers in the surrounding media cause significant problems when characterizing high-voltage SiC devices. SF₆ was found to a better ambient in this respect than fluids such as Fluorinert FC-70 [6] and oils used for isolating transformers and capacitors.

The reverse leakage current is very low, and indeed too low to be easily measured: up to the breakdown voltage the reverse leakage current is less than the measurement limit of 1 μA at all temperatures tried, i.e. up to 200 °C, and it is also less than the corresponding limit of 10 nA up to 100 V over the same temperature range.

The forward current–voltage (I – V) characteristics are shown in Fig. 2. As can be seen, the voltage drop is as low as 3.0 V at 300 A cm⁻²; higher current density cannot be reached in these 200 μm diameter diodes, as the current is maximized to 100 mA by the instrumentation. At low current densities, a moderate excess recombination is observed.

4. Comparison with simulation

The minority carrier lifetime was derived from measurements of the decline in electroluminescence when abruptly decreasing the forward current, and found to be 25 ns at room temperature; no great variation with the injection level was observed. This value was further corroborated by reverse recovery studies that yielded lifetimes in the range 25–40 ns.

Using a recently developed version of TMAs Medici [7] device simulator that treats anisotropic materials such as 6H SiC correctly, we have simulated the forward I – V characteristics of the experimental structures. As seen in Fig. 2,

the use of 25 ns for room temperature minority carrier lifetime yields good agreement with experimental I – V curves.

The excess recombination in the low current regime is underestimated by the simulator, as there are for example no interface traps specified. The simulator overestimates the effects of series resistance in the substrate, since the actual current spreading is larger than allowed by the simulated geometry.

Figure 3 shows the simulated carrier concentrations across the pin diode structure at room temperature and 500 A cm⁻², using the measured carrier lifetime. For the case “1”, it can be seen that the carrier concentration in the slowly doped n⁻ base is three times the doping level. There is thus an injected plasma and a significantly improved conductivity. It is interesting to note that this injection level is reached in spite of the very low efficiency of the p⁺ emitter ($\gamma = 0.4$). This low value is partly caused by the low activation of the p-type dopant, aluminium, that has an activation energy as high as 220 mV; if the emitter efficiency is calculated while neglecting the incomplete ionization of the acceptor, the value becomes $\gamma = 0.6$. Although this is also a low value, the injected carrier concentration becomes as much as three times higher than when the incomplete ionization is properly accounted for (see Fig. 3).

Too short carrier lifetime in the base layer cause unacceptably high voltage drops in pin diodes. While 25 ns was found to be a sufficient lifetime for the above device, which has a calculated bulk breakdown voltage of 1700 V, the question arises about the shortest lifetime necessary to avoid high voltage drop for this and higher voltage designs. To find out, a set of simulations were carried out for diodes with calculated bulk breakdown voltages of 1.7 kV, 3.6 kV and 5.4 kV. The 1.7 kV diode has a layer structure identical to the experimental diode. The 3.6 kV diode has a 1 μm p⁺ layer doped to $1 \times 10^{19} \text{ cm}^{-3}$ on top of a 20 μm n⁻ layer doped to $5 \times 10^{15} \text{ cm}^{-3}$, while the 5.4 kV diode has a 1 μm p⁺ layer doped to $1 \times 10^{19} \text{ cm}^{-3}$ on top of a 30 μm n⁻ layer doped to $2 \times 10^{15} \text{ cm}^{-3}$. In both structures, there is a 0.5 μm n⁺⁺ buffer layer included between the n⁻ base and the n⁺ substrate, similar to the experimental case.

In Fig. 4, the forward voltage drop at a current density of 500 A cm⁻² is plotted as a function of minority carrier life-

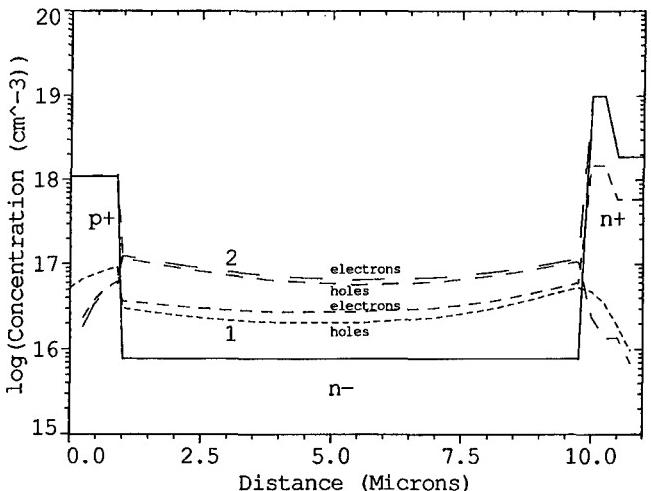


Fig. 3. Simulated electron and hole concentration for the experimentally studied structure at room temperature and 500 A cm⁻². The curves labelled “1” are calculated while including an incomplete ionization term in the simulation; the curves “2” are calculated neglecting this effect.

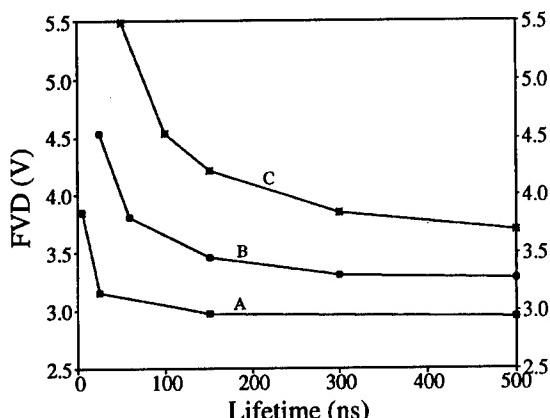


Fig. 4. Forward voltage drop FVD as a function of carrier lifetime in the base for pin-diodes with different bulk breakdown voltages, (A) 1700 V, (B) 3600 V, (C) 5400 V.

time. As can be seen, the carrier lifetime in our present diodes should not be substantially lower than the actual 25 ns, or a forward voltage drop penalty results. The 3.6 kV design, which in practical terms can be regarded as a 3.0 kV design, assuming reasonably good junction termination, requires a lifetime of about 100 ns to avoid an excessive voltage drop. Correspondingly, the 5.4 kV bulk breakdown design – aimed as a 4.5 kV device – requires close to 200 ns lifetime to reach an acceptable voltage drop. When evaluating the viability of high-voltage SiC pin-diodes, these figures should be compared with the longest lifetime reported to date (to the authors' knowledge) 200 ns [8].

5. Conclusions

We conclude that silicon carbide pin rectifiers can have low forward voltage drops even when designed for blocking

voltages of several kilovolts, owing to the conductivity modulation caused by injected carriers. The carrier lifetimes required are not overly large, and should be attainable, considering the advancement of SiC epitaxy; currently, lifetimes attained suffice for 3–4 kV diodes.

In modelling the properties of SiC bipolar devices, we conclude that the incomplete ionization must be accounted for, or the emitter efficiency of p-type emitters will be heavily overestimated.

Finally, regarding the measurement of the breakdown properties of unpassivated SiC devices that sustain more than 600 V, which is difficult due to destructive flashovers in the surrounding media, we conclude that gaseous SF₆ offers a more useful ambient than any liquid tried.

Acknowledgements

We are indebted to Dr. Peder Bergman at the University of Linköping for the electroluminescence measurements.

References

1. For an overview of SiC technology, see P. A. Ivanov and V. E. Chelnokov, *Semicond. Sci. Technol.* **7** 863–880 (1992).
2. Bhatnagar, M. and Baliga, B. J., *IEEE Trans. Electr. Dev.*, **40**, (3), 645–655 (1993).
3. Sze, S. M., "Physics of Semiconductor Physics" (J. Wiley & Sons 1981), 2nd Edn., p. 207.
4. Durham, NC, U.S.A.
5. Neudeck, P. G. *et al.*, *Appl. Phys. Lett.* **64**, (11), 1386–1388 (1994).
6. A viscous liquid with high dielectric strength marked by the 3M corporation.
7. Medici is a drift-diffusion based FEM device simulation programme supplied by Technology Modelling Associates Inc, Palo Alto, CA, U.S.A.
8. Saddow, S. E. *et al.*, SPIE vol. 1873, Optically Activated Switching III, p. 110 (1993).

Metallizations of InP Based on Transition Metals

T. Clausen

Physics Department, Building 309, Technical University of Denmark, 2800 Lyngby, Denmark

and

O. Leistikko

Mikroelektronik Centret, Building 345E, Technical University of Denmark, 2800 Lyngby, Denmark

Received May 2, 1994; accepted in revised form June 27, 1994

Abstract

Single element Ni and Cr metallizations to low doped n- and p-InP have been investigated to clarify how they affect electrical performance of advanced multilayer metallizations for ohmic contact and Schottky diode applications. It was found that the Ni and Cr contacts to n-InP show unstable ohmic behaviour at all annealing temperatures (300–500 °C), and that Ni and Cr metallizations to p-InP show diode behaviour. The electrical behaviour of the contacts were compared to phase formation paths of Ni-In-P and Cr-In-P ternary systems. From this it was found that amorphous ternary phases form first by interdiffusion of Cr and Ni into the InP crystal. Subsequently, at higher annealing temperatures, crystallization and eventually phase separation determines the electrical properties of the contacts. After annealing at 500 °C the Cr diodes to p-InP are almost ideal, as deduced from barrier height measurements using a combination of I-V and C-V methods. This indicates that the metallization forms a two-layer structure with Cr-P phases lying above pure In. The Ni diodes to p-InP annealed at 500 °C, on the other hand, are not ideal, indicating that Ni-In and Ni-(In)-P phases in parallel are determining the electrical properties of the diode.

1. Introduction

Ni and Cr are commonly used in multilayer metallizations to III-V semiconductors, a classical example being the AuGeNi ohmic contact metallization to n-type materials [1–3]. For this particular application it is found that the addition of Ni is beneficial in that low values of the specific contact resistance, r_c , result over a broader range of annealing temperatures [1–3]. We have previously published results on AuGeX and AuX (X = Ni or Cr) metallizations to n-InP [1, 4]. Briefly, we found that r_c is very much affected by the growth of Ni-P/Cr-P phases in contact with the semiconductor [1, 4]. In this paper we report on the electrical properties of annealed Ni and Cr metallizations to n- and p-InP to further identify the contact mechanisms.

2. Experimental

Ni (100 nm) and Cr (100 nm) were deposited by r.f. sputter deposition and e-beam evaporation on n- and p-InP samples ($15 \times 15 \text{ mm}^2$) sliced from 2-inch wafers that had been covered with SiO_2 and prepattered before metal deposition [1, 4]. N-type InP wafers were undoped ($N_a = 10^{15} \text{ cm}^{-3}$) and p-type InP wafers were Zn-doped ($N_a = 5 \times 10^{16} \text{ cm}^{-3}$). AuGe and AuZn alloys were used on the back of the samples to form ohmic contacts to N- and P-InP. Annealing was performed in a RC 2400 Polaron alloying furnace using a forming gas atmosphere (cracked

ammonia, 10% H_2 in N_2). The annealing time was 2 minutes (heating rate: 100 °C/min).

For electrical characterization of the contacts the samples were placed on an heatable aluminum chuck in a probing station. Temperature (T) measurements were done directly underneath the samples by mounting a standard thermocouple. Electrical characterization included I-V, I-V-T and C-V measurements for the contacts.

3. Results

The results of the I-V characterization for Ni- and Cr/p-InP contacts fabricated on the same samples are shown in Fig. 1. As can be seen from this figure, diode behaviour is observed for all the annealing temperatures investigated. Also, the I-V behaviour as a function of annealing temperature is the same when comparing Ni and Cr diodes, apart from large non-idealities for reverse and low forward voltages for the Ni diodes. The as-deposited diodes, and diodes annealed at 300 °C have large series resistances in comparison with contacts annealed above 300 °C, implying that the metallization resistance changes. Furthermore, the current through both the Cr and the Ni diodes are limited solely by recombination. The diodes again improve with respect to transport properties after annealing at temperatures above 300 °C. At 500 °C the Cr diodes have a soft temperature independent breakdown at $\sim -0.3 \text{ V}$.

The C-V plots for the Ni and Cr diodes, shown in Fig. 2, are very complex. Only for annealing temperatures of 500 °C the slope of the lines corresponds to the background doping density level of the low doped p-InP. For annealing temperatures below 500 °C the slope of the lines does not correspond to the background doping density level; the as-deposited diodes and the diodes annealed at 300 °C all have non-linear $1/C^2-V$ characteristics. Built-in voltages, V_{bi} , for the diodes are measured to be 1.17 V for the Ni diodes and 0.79 eV for the Cr diodes annealed at 500 °C.

For Cr diodes annealed at 500 °C, the ideality factor was found to be 1.1, indicating that an almost ideal diode has formed with a barrier height (assuming thermionic emission with $A^* = 60 \text{ A/cm}^2 \text{ K}^2$ [5]) of 0.9 eV. This value is in good agreement with V_{bi} found from C-V measurements if the potential difference between the Fermi level and the valence band ($\sim 0.12 \text{ eV}$) is added. The barrier height for the Ni diodes annealed at 500 °C is difficult to extract from the I-V plots because of the non-idealities. Using the ideality factor

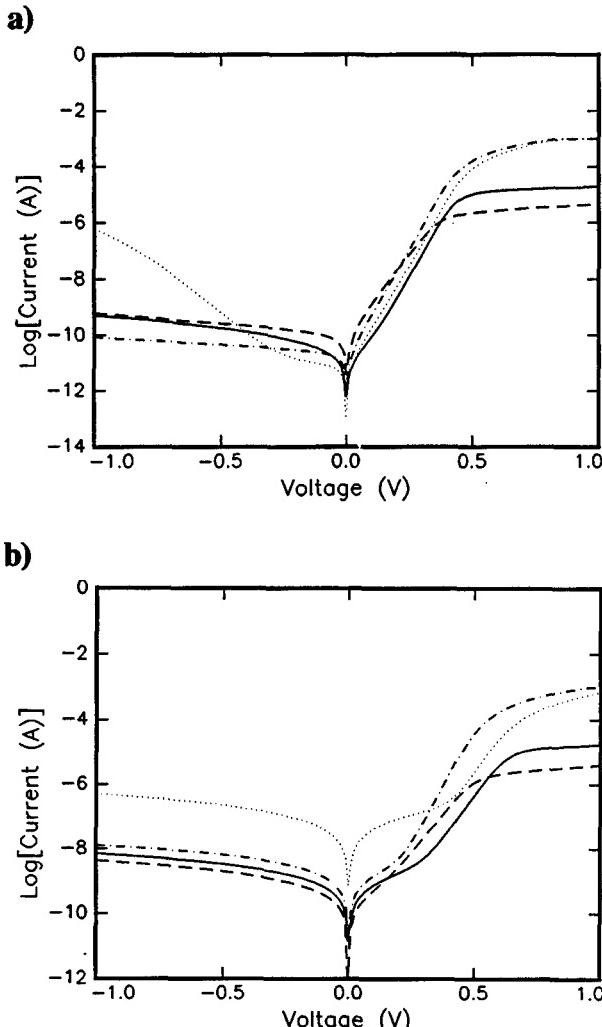


Fig. 1. $I-V$ plots for (a) Cr/p-InP and (b) Ni/p-InP diodes annealed at different temperatures: solid lines; as-dep., dashed lines; 300 °C, dashed-dotted lines; 400 °C and dotted lines; 500 °C.

for the Cr diodes annealed at 500 °C, an upper limit for the barrier height was calculated by simulation to be 1.06 eV. Thus, there is a significant difference in the values of the barrier height for the Ni diodes depending on the characterization method ($I-V-C-V$).

For Ni and Cr metallizations to n-InP (with patterned Au/Cr probing pads) ohmic behaviour was observed for annealing temperatures of 300 °C, 400 °C and 500 °C. However, the stability of the ohmic contacts was poor, with the most stable contacts being annealed at 400 °C. For as-deposited contacts we found that the Cr contacts showed diode behaviour with a barrier height of 0.5 eV, while the Ni contacts were all ohmic. This is shown in Fig. 3, where a method based on measurements of r_c under different ambient temperatures are used to determine the effective Schottky barrier height from the slope of the line in a semi-logarithmic $r_c T$ vs. $1/T$ plot, assuming that thermionic emission is limiting the current flow [1, 4]. From Fig. 3, it is seen that the change from diode to ohmic behaviour at 400 °C can be explained by a lowering of the effective Schottky barrier height from ~0.5 eV to ~0.1 eV.

For Ni based contacts the interpretation of the plots is somewhat more difficult, in that the slope in the semilogarithmic $r_c T$ vs. $1/T$ plot for the as-deposited contacts is negative (Fig. 3). Also, after annealing at 400 °C, two slopes

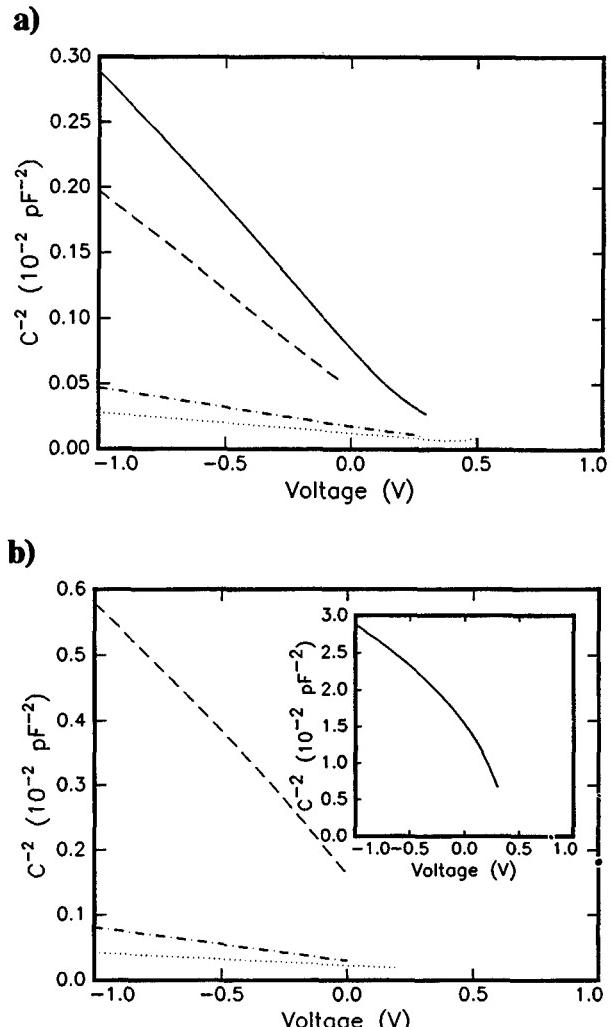


Fig. 2. $C-V$ plots for (a) Cr/p-InP and (b) Ni/p-InP diodes annealed at different temperatures: solid lines; as-dep., dashed lines; 300 °C, dashed-dotted lines; 400 °C and dotted lines; 500 °C.

are deduced from Fig. 3; a large slope at temperatures slightly above room temperature and a smaller slope at higher temperatures. The negative slope for the as-deposited contacts does not correspond to a negative Schottky barrier. Instead the current flow in the contacts is limited by drift

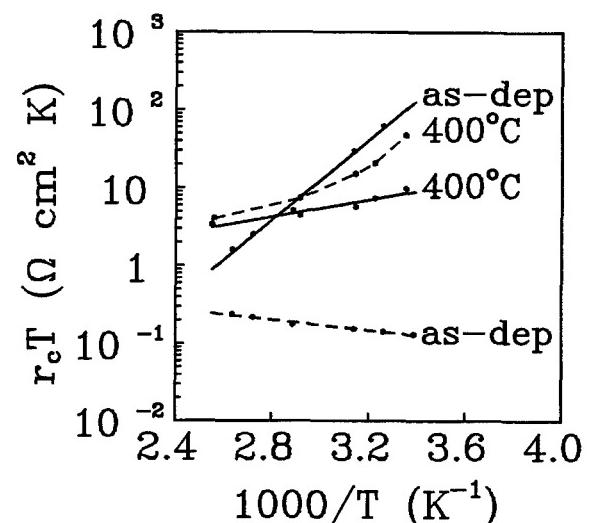


Fig. 3. Temperature activation plots for Cr/n-InP ohmic contacts (solid lines) and Ni/n-InP ohmic contacts (dashed lines).

and diffusion across the potential difference between the Fermi level and the conduction band (E_c) established in a Schottky contact with an effective barrier height close to 0 eV. This was shown also to be the case for both AuNi and AuGeNi contacts to n-InP [1, 4].

4. Discussion

The results for the Ni and Cr metallizations to InP agree with our previous findings; the transition metals are mainly responsible for the observed high/low specific contact resistance to InP depending on the bulk carrier type [1, 4]. For n-InP they form current paths between the alloyed metallization and the bulk by lowering the effective Schottky barrier height. For p-InP they form blocking parts in the alloyed diode, so that barrier lowering phases only form in relatively small amounts [1, 4].

The major problem in the Cr contact experiments has been the absence of an element for In incorporation (e.g. Au, in the alloyed Au-based metallizations). No Cr-In binary phase can exist, but Cr-P phases can exist in thermodynamical equilibrium with InP and In [6]. We believe, from previous findings [1, 4] and work done by Ivey *et al.* on AuCr contacts to InP [6], that the annealed Cr/InP contacts form a two-layer structure at their end point of reaction (i.e. 500 °C) with Cr-P lying above a metallic In layer that is in close contact with the InP. The formation of a single phase towards the InP is supported by the observation of a nearly ideal diode structure at 500 °C annealing. The temperature independent breakdown at reverse voltages can be explained in terms of regrown heavily doped InP layers forming in a melting/regrowth sequence caused by the pure In ($T_{melt} = 156$ °C) at the InP surface.

A multiplicity of Ni-In and Ni-P binary phases can coexist in thermodynamical equilibrium with InP [7]. However, ternary amorphous and crystalline phases are more commonly found in the Ni-In-P system [8, 9]. The amorphous ternary phase is observed for mild annealing conditions (250 °C/5 min [9], 250 °C/15 min [8]); we believe that this amorphous phase is responsible for the observed electrical properties of both the as-deposited contacts and the contacts annealed at 300 °C. The as-deposited contacts received some process induced postdeposition annealing, below 200 °C, which induced some reaction processes. During annealing Ni and Cr (as the observed electrical properties are identical to those for Ni) are believed to diffuse into the bulk InP compensating shallow donors and initiating the amorphous phase formation between the three constituents. This might explain the anomalies in the C-V plots and the observed high series resistances as amorphous phases tend to have a higher resistance than crystalline ones.

After annealing at 400 °C and 500 °C, the diode characteristics improved and the ohmic contacts were more stable (400 °C). Both Sands *et al.* and Mohney *et al.* have established that the thermodynamically favored end point in the reaction between Ni and InP is a ternary monoclinic Ni₂InP phase [8, 9]. We believe that the growth of this phase determines the electrical properties of the Ni contacts for annealing temperatures above 300 °C with perhaps binary phases forming in small amounts at 400 °C. Thus, the difference between the Cr and the Ni contacts is due to the ability for Ni to form ternary crystalline phases, while for Cr only stable Cr-P phases exist. The large recombination currents observed for the Ni diodes are probably due to the poor epitaxial relationship between the monoclinic Ni₂InP ($a = 0.6795$ nm, $b = 0.5269$ nm, $c = 0.6436$ nm, and $\beta = 94.73^\circ$ [9]) phase and the InP. During annealing at temperatures above 300 °C, phosphorous loss from the reacting region of the contact can be a serious problem, but we have tried to minimize these losses by annealing in semiclosed systems with small volumes.

5. Conclusion

An identical behaviour with respect to the electrical properties have been demonstrated for Ni and Cr metallizations to n- and p-InP for annealing temperatures up to 300 °C, probably due to the formation of amorphous ternary phases. For annealing temperatures about 300 °C thermodynamically favored crystalline ternary phases form for the Ni contacts, but for the Cr contacts phase separation into elemental In and Cr-P stable phases occur. For Ni contacts this makes large recombination currents flow in the contacts because of the large mismatch in lattice parameters between Ni₂InP and InP. For the Cr contacts, with In in close contact with InP and Cr-P lying above, an almost ideal contact is formed.

References

1. Clausen, T. and Leistikko, O., *Thin Solid Films* **232**, 215 (1993).
2. Murakami, M., *Materials Science Reports* **5**, 273 (1990).
3. Del Alamo, J. A. and Mizutani, T., *Solid-State Electronics* **31**, 1635 (1988).
4. Clausen, T. and Leistikko, O., *Applied Surface Science* **74**, 287 (1994).
5. Robinson, G. Y., in: "Physics and Chemistry of III-V Compound Semiconductor Interfaces" (Edited by C. Wilmsen) (Plenum Press, New York 1985), p. 73.
6. Ivey, D. G., Bruce R. and Piercy G. R., *Solid-State Electronics* **31**, 1251 (1988).
7. Clausen, T., Ph.D. thesis, Technical University of Denmark (1993).
8. Sands, T. *et al.*, *Appl. Phys. Lett.* **50**, 1346 (1987).
9. Mohney, S. E. and Chang, Y. A., presented at the 1994 MRS Spring Meeting, to be published in MRS symposium proceedings vol. 337.

Impedance Field and Microwave Power Generation in InP Diodes

V. Gružinskis, E. Starikov and P. Shiktorov

Semiconductor Physics Institute, A. Goštauto 11, 2600 Vilnius, Lithuania

Received April 28 1994; accepted in revised form July 12, 1994

Abstract

A closed hydrodynamic (HD) approach is used to carry out a comprehensive spatial analysis of the dynamic features of submicron n^+nn^+ InP diode both in time and frequency domains. The contribution of each part of the device, when operating as microwave power generator, is analyzed through the spatial profiles of the impedance-field spectrum. The usual subdivision of the n -region into a passive (dead-zone) and active zone is carried out. The dead zone is found to manifest itself as a purely real resistance which is practically independent of the frequency. One or more spatial zones which are responsible for the generation are shown to be formed in the active region of the diode. By reducing the length of the n -region, under the condition that the total current is constant in time, the additivity of the contributions from each part of the device into the generation spectrum is demonstrated.

1. Introduction

The near-micron n^+nn^+ InP diodes are widely used in modern semiconductor generators of short millimeter waves [1–6]. To improve the high-frequency performance of these generators various doping profiles [6, 7] and a reduction of the n -region length [8] are suggested. For a proper choice of the diode length, doping profile, etc., it is firstly necessary to investigate theoretically, how a variation of these parameters influences on the main characteristics of the device. To account for the main physical processes inherent in such diodes (the nonlocal heating of carriers, the velocity overshoot, the intervalley transfer, the carrier concentration and self-consistent electric field redistribution during transit-time dynamics, etc.) the theoretical analysis is performed in the framework of either kinetic [5, 6, 7, 9] or hydrodynamic [8, 10–12] approaches coupled with the Poisson equation. In so doing, a quantitative analysis of the parameters which allow for a spatial analysis of various physical quantities has to be preferred. Indeed, by allowing one to construct a spatial map of the device properties of interest, the designing of the device is significantly facilitated. The main aim of this paper is to demonstrate that the impedance field can be successfully used for this sake providing an important physical information both in time and frequency domains.

2. The impedance field approach

By definition, in the frequency domain the local impedance-field, $\nabla Z(\omega, x)$ relates the linear response of the electric field $\delta E_\omega(x)$ to a small harmonic perturbation of the total current $\delta j_\omega(x)$ at circular frequency $\omega = 2\pi f$ in point x as:

$$\delta E_\omega(x) = \nabla Z(\omega, x) \delta j_\omega(x) \quad (1)$$

For the one-dimensional structure considered here, the total current is constant in space, and the integration of eq. (1) over the whole diode gives:

$$\delta U_\omega = Z(\omega) \delta j_\omega \quad (2)$$

Here δU_ω is the linear response of the terminal voltage caused by a harmonic perturbation of the total current and $Z(\omega) = \int_0^l \nabla Z(\omega, x) dx$ (where l is the diode length) is the small-signal impedance of the whole diode. To simulate the carrier transport in n^+nn^+ -InP diodes the closed hydrodynamic approach [13–15] based on the carrier number, drift velocity and mean energy conservation equations coupled with the Poisson equation for the self-consistent electric field has been used. This model was demonstrated to give the results which are in an excellent agreement with the Monte Carlo calculations for both homogeneous semiconductors [13–15] and short n^+nn^+ structures [8, 9]. To calculate the impedance field, we use an impulsive procedure which enables us to obtain simultaneously the spectra of both $\nabla Z(\omega, x)$ and, hence, $Z(\omega)$ in the frequency range of interest. To this end, let us consider the stationary-state of the diode as characterized by a potential drop U_{d0} and a total current-density j_0 . Then, at a given initial time $t = 0$ an impulsive delta-like perturbation of the total current density, δj_0 , is introduced leading to an initial perturbation of the voltage drop in the form $\delta U_d(0) = \delta j_0 / \epsilon_r \epsilon_0$. Here ϵ_0 is the vacuum permittivity, ϵ_r the relative static dielectric-constant of the material. The voltage perturbation leads to a homogeneous perturbation of the electric field inside the diode $\delta E_0 = \delta U_d(0)/l$ which, in turn, causes a time variation of a conduction current. As a result, $U_d(t)$ also begins to change finally relaxing to the initial value U_{d0} , since all perturbations in velocity, concentration, and voltage vanish in time under constant current-operation.

To investigate the spatial profile of the voltage response, we evaluate the response function of the local electric-field in each point of the device, $D_E(t, x)$, defined as:

$$D_E(t, x) = \frac{1}{\epsilon_r \epsilon_0} \frac{\delta E(t, x)}{\delta E_0} \quad (3)$$

Integration of $D_E(t)$ throughout the whole device gives the voltage-response function $D_U(t)$. By Fourier transforming eq. (3), one obtains the impedance field, $\nabla Z(f, x)$, at frequency f which is defined as:

$$\nabla Z(f, x) = \int_0^\infty D_E(t, x) \exp(-i\omega t) dt \quad (4)$$

3. Results and discussion

We consider a n^+nn^+ structure at $T = 300\text{ K}$ with parameters which are typical of to-date diode generators [1–6]: the n -region length $l_n = 1\text{ }\mu\text{m}$, $n = 2 \times 10^{16}\text{ cm}^{-3}$, $n^+ = 10^{18}\text{ cm}^{-3}$. Abrupt homojunctions between n and n^+ regions are assumed. Let us start from the near-equilibrium conditions. Figure 1 shows the time dependencies of the local electric field response functions calculated at $U_d = 0$ for two points of the diode chosen from the n and n^+ (solid and dashed curves, respectively). The damped oscillations are caused by the plasma processes in these regions which give separate contributions into the time dependence of the voltage response function of the whole diode shown in Fig. 2 for two diodes with the same length of the n region but with various lengths of the n^+ contacts: 0.1 and 0.3 μm (solid line), and 0.05 and 0.1 μm (dotted line), for the cathode and anode contacts, respectively. As it follows from Fig. 2 the plasma oscillations corresponded to the n^+ regions are more pronounced for the structure with the longer contacts. Figure 3 presents the real and imaginary parts of the small-signal impedance.

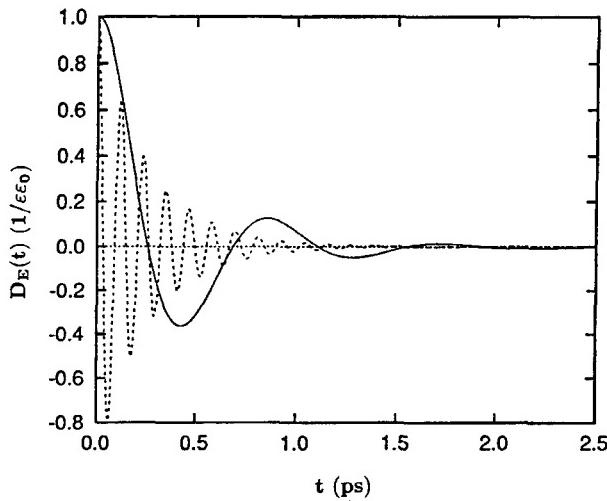


Fig. 1. Normalized response functions of the local electric field as a function of time for $U_d^0 = 0$ under current driven operation at two different points inside the n and n^+ regions of the diode (solid and dashed lines, respectively). Diode parameters are: $n^+ = 10^{18}\text{ cm}^{-3}$, $n = 2 \times 10^{16}\text{ cm}^{-3}$, $l_1^+ = 0.1\text{ }\mu\text{m}$, $l_n = 1.0\text{ }\mu\text{m}$, $l_2^+ = 0.3\text{ }\mu\text{m}$.

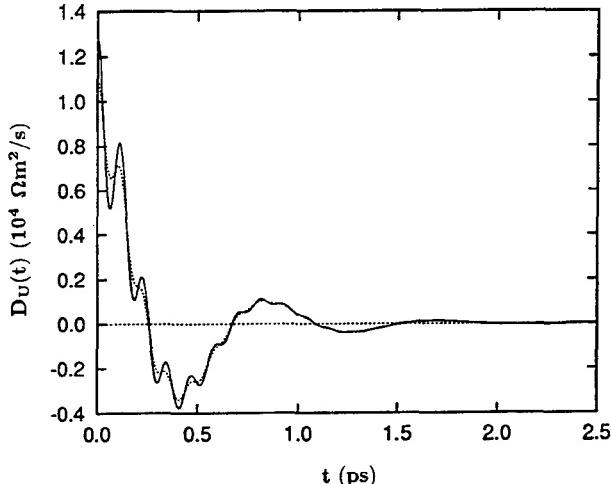


Fig. 2. Voltage response-function as a function of time calculated under current-driven operation for the diode of Fig. 1 (solid line) and with reduced contact lengths $l_1^+ = 0.05\text{ }\mu\text{m}$, $l_2^+ = 0.1\text{ }\mu\text{m}$ (dotted line). $U_d^0 = 0$.

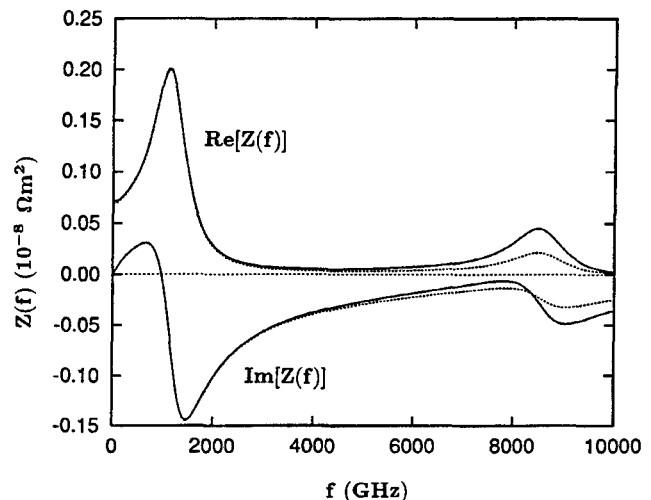


Fig. 3. Frequency dependence of the real and imaginary part of the small-signal impedance, $Z(f)$. The notation is the same of Fig. 2.

signal impedance of the whole diode, $\text{Re}[Z(f)]$ and $\text{Im}[Z(f)]$, respectively, obtained by Fourier transformation of the voltage response functions presented in Fig. 2. The plasma oscillations in n and n^+ regions results in two peaks of the $\text{Re}[Z(f)]$ which position independes of the structure lengths and corresponds to the plasma frequency in these regions only. Thus, at the thermal equilibrium the diode response has a local character, and processes in various regions can be considered independently one from another. In the following, in order to obtain quasi thermal-equilibrium conditions for carrier transport at the contact ends, the cathode and anode n^+ -region lengths are taken to be 0.1 and 0.2 μm , respectively.

To analyze the high-frequency performance of short n^+nn^+ InP diodes under biasing conditions, one should recall that a considerable role is played by drift-velocity overshoot, which increases the difference between the maximum and minimum drift velocities inside the n region, and strengthens the usual Gunn effect [10]. This is illustrated in Fig. 4, which present a stationary profiles of the drift velocity in the structures with different lengths of the n region calculated for the same total current j_0 . For the case

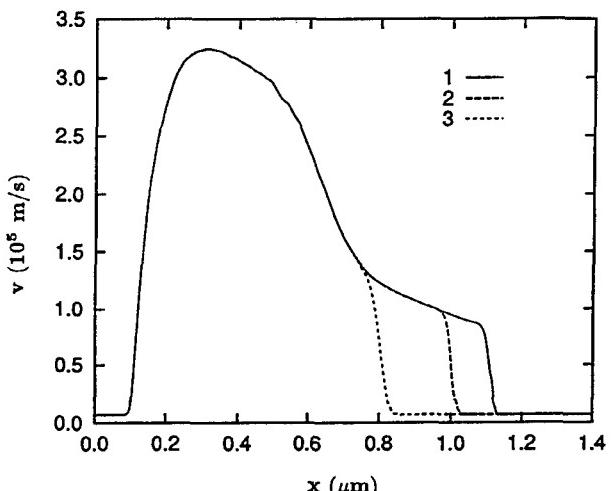


Fig. 4. Spatial profiles of the drift velocity with different n -region length l_n : 1–1.0 μm , 2–0.9 μm , 3–0.72 μm (curves 1 to 3, respectively). All curves are obtained at the same current density $j_0 = 1.07 \times 10^9 \text{ A/m}^2$ (for curve 1 this corresponds to $U_d = 8\text{ V}$).

of $l_n = 1 \mu\text{m}$ this corresponds to $U_d = 8 \text{ V}$. The velocity overshoot results in the spatial negative differential-conductivity (SND) in the space region $0.3 < x < 1.1 \mu\text{m}$, where the drift velocity decreases with increasing the spatial coordinate. It is an active zone of the diode where perturbations of the carrier concentration can grow in time. Figure 5 reports the voltage response function at increasing values of the applied voltage. The general shape of each curve is similar and it can be analyzed in terms of a short and long time behavior. At short times we find a sharp decrease of $D_U(t)$ which is modulated by the plasma oscillations associated with the n^+ -regions. (By comparison with Figs 1 and 2 one can conclude that the plasma oscillations are caused by the cold n^+ -regions, only). At long times we find a bell-shape behavior. To discuss its origin Fig. 6 shows the time evolution of the perturbed carrier-concentration, $\delta n(t, z) = n(t, z) - n_0(z)$, $n_0(z)$ being the stationary profile corresponding to $U_d^0 = 8 \text{ V}$). The initial perturbation of the applied voltage and, hence, of the electric field inside the

diode, leads to the appearance of a space-charge wave which is located in the active region of the diode (curve 1, $t = 1 \text{ ps}$). Then, this wave turns in an accumulation layer followed by a narrow depletion region near the anode (curve 2, $t = 2 \text{ ps}$). By comparing with Fig. 4, one can conclude that the accumulation layer is formed in the region of the diode where the negative slope of the drift-velocity as a function of space is maximum. After its formation, which takes about 2 ps, the accumulation layer begins to propagate towards the anode. The layer disappears reaching the anode with a negative value of δn . Figure 7 shows the time dependence of the response function of the local electric field calculated for $U_d^0 = 8 \text{ V}$ at several points of the diode with coordinates, as measured from the cathode terminal, respectively of $x = 0.05, 0.3, 0.65, 1.05 \mu\text{m}$. The first point is placed inside the cathode. Here the response function exhibits a plasma-oscillation pattern which is damped by the collision time. The same pattern is found inside the anode. The second point corresponds in space to the maximum value of drift-velocity overshoot. Here the response function is found to decay nearly exponentially on the time scale of momentum relaxation. For the other points, which are placed between the maximum value of the drift-velocity overshoot and the anode, the response function exhibits a pronounced bell-shaped tail. The maximum value of the bell-shape and its corresponding time is found to increase by increasing the coordinate along the diode up to the point where the carrier mean-energy has its highest value (curve 4 in Fig. 7). Then the maximum of the bell-shape tail begins to decrease rapidly and vanishes at the beginning of the anode region. Thus, the bell-shape behavior of $D_E(t, x)$, is caused by the propagation of the accumulation layer across the diode and is more pronounced the higher is the applied voltage. When the propagation of the accumulation layer is terminated, the voltage response function vanishes.

The bell-shape behavior of the voltage response function is responsible for the appearance of one or more minima in the frequency spectrum of the real part of the small-signal impedance of the diode, $\text{Re} [Z_d(f)]$. This is illustrated in Fig. 8, where the spectra of $\text{Re} [Z_d]$ are presented for $U_d^0 = 1, 3, 5$ and 8 V (curves 1 to 4, respectively). In the frequency

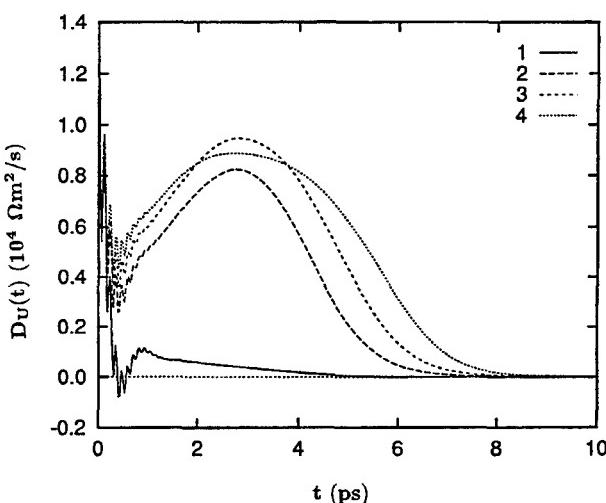


Fig. 5. Voltage response-function as a function of time calculated under current-driven operation for the diode of Fig. 1. Curves 1 to 4 corresponds to $U_d = 1, 3, 5$ and 8 V .

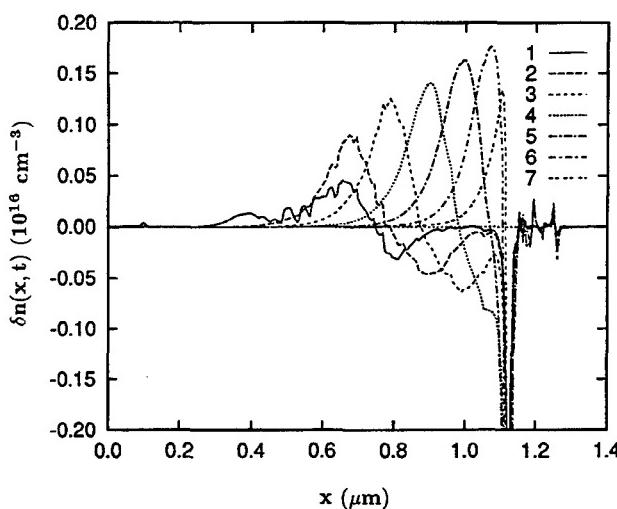


Fig. 6. Concentration perturbations $\delta n(t, x) = n(t, x) - n_0(x)$ with respect to the stationary concentration profile $n_0(x)$ as a function of the coordinate x along the n^+nn^+ diode calculated under the current driven operation at various times $t = 1, 2, 3, 4, 5, 6$ and 7 ps after the voltage perturbation (curves 1 to 7, respectively). $U_d^0 = 8 \text{ V}$, $\delta U_0 = 0.05 \text{ V}$.

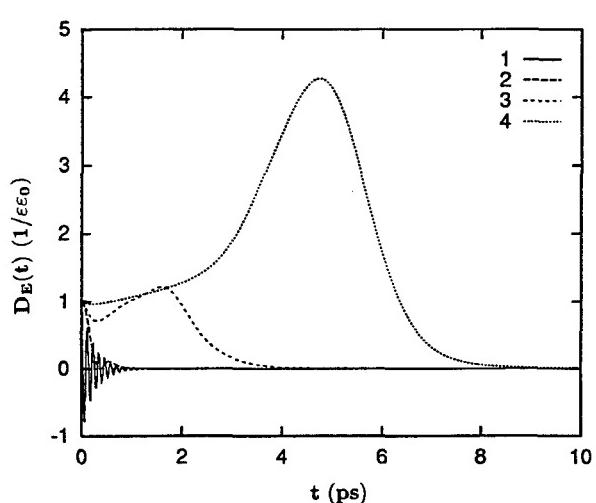


Fig. 7. Normalized response functions of the local electric field as a function of time for $U_d^0 = 8 \text{ V}$ under current driven operation at different points inside the diode of Fig. 1 as measured from the cathode terminal: $z = 0.05, 0.3, 0.65$ and $1.03 \mu\text{m}$ (curves 1 to 4, respectively).

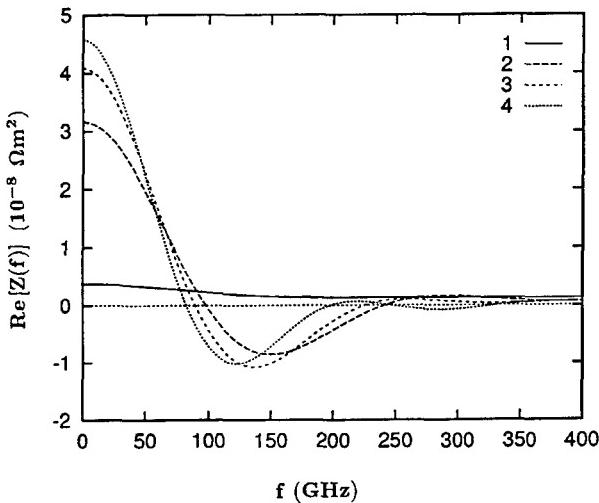


Fig. 8. Frequency dependence of the real part of the small-signal impedance calculated from the time-dependent voltage response functions presented in Fig. 5. The notation is the same of Fig. 5.

range where $\text{Re}[Z_d(f)] < 0$ an amplification and generation of the microwave power is possible. Figure 9 summarizes the generation frequency tuning (horizontal lines) obtained experimentally in [1–6] together with the $\text{Re}[Z(f)]$ spectrum calculated by the HD approach at $U_d = 5 \text{ V}$ (solid line). One can see, that all available experimental data practically fully cover the frequency region of the generation predicted by the linear theory.

Let us discuss now the contributions of the various parts of the diode to $Z(f)$ and its dependence on reducing the diode length. The real part of the small-signal impedance, $\text{Re}[Z(f)]$, calculated for the considered n^+nn^+ structure is reported in Fig. 10 for different lengths of the n -region. For the case of $l_n = 1 \mu\text{m}$ (see curve 1), the amplification condition $\text{Re}[Z(f)] < 0$, is fulfilled inside the two bands: $f = 70 \div 200$ and $250 \div 340 \text{ GHz}$ where microwave power generation is possible. We remark that, by shortening the n -region, the condition for amplification shifts to high frequencies, as expected. Figures 11 (a) and (b) report the spatial profile respectively of the real and imaginary parts of

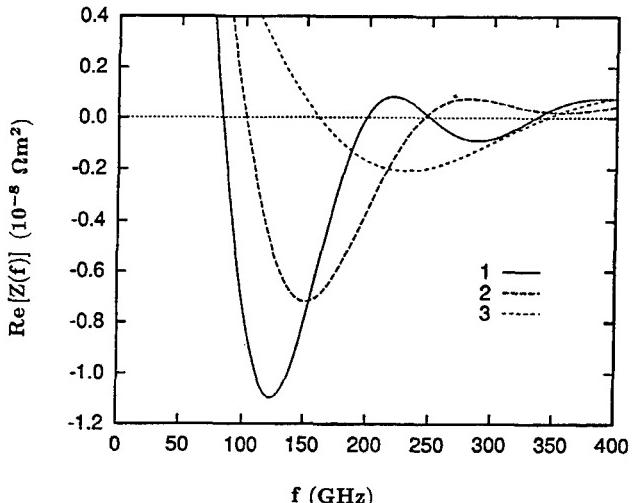


Fig. 10. Frequency dependence of the real part of the small-signal impedance calculated for diodes with different n -region length l_n . The notation is the same of Fig. 4.

the impedance field for the structure with $l_n = 1 \mu\text{m}$. Curves 1, 2 and 3 correspond, respectively, to the frequencies $f_1 = 55 \text{ GHz}$, $f_2 = 125 \text{ GHz}$ and $f_3 = 290 \text{ GHz}$. We recall that $\text{Re}[Z(f)]$ is always positive for f_1 and reaches a minimum

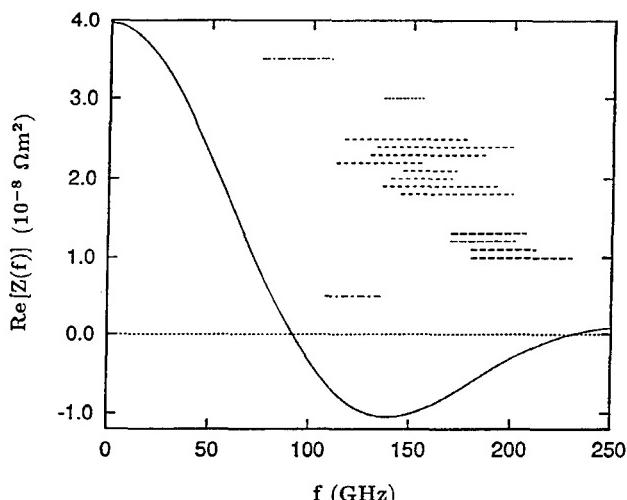
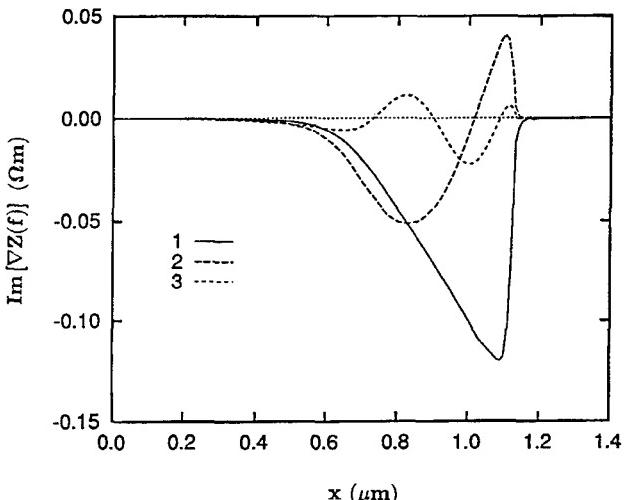
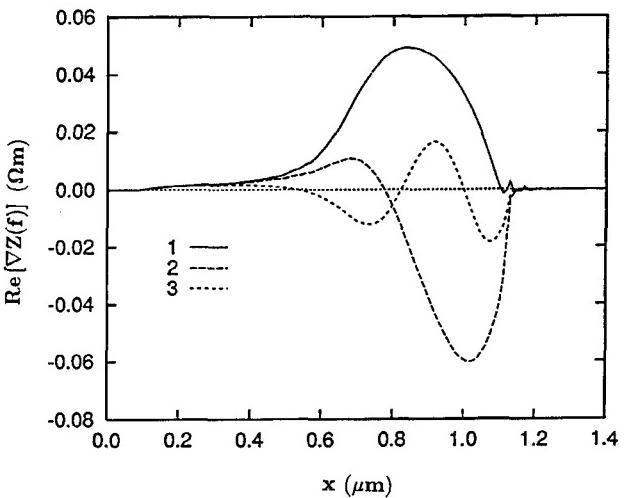


Fig. 9. Frequency dependence of the real part of the small-signal impedance (solid line) calculated with the HD approach at $U_d = 5 \text{ V}$ for the same diode of Fig. 1. Horizontal lines and separate groups of lines summarize the experimentally observed tuning of the generation frequency, and correspond from top to bottom, respectively to Refs [1–4, 6].

Fig. 11. Spatial profiles of (a) the real and (b) imaginary part of the impedance field calculated with the HD approach for the same diode of Fig. 1 at three different values of the frequency: 1–55 GHz, 2–125 GHz, 3–290 GHz, $l_n = 1.0 \mu\text{m}$. $U_d = 8 \text{ V}$.

in the first and second generation bands for f_2 and f_3 , respectively (see Fig. 10). As it follows from Fig. 11, $\text{Re}[\nabla Z(f)]$ is practically independent from frequency in the near-cathode area of the n -region ($x = 0.1 \div 0.3 \mu\text{m}$). From the comparison with curve 1 of Fig. 4, we find that this is the region where the drift velocity exhibits a sharp increase up to its maximum value. This space region is usually called as the dead zone of short diodes. Comparing Figs 4 and 11, one can conclude that the dead zone manifest itself as a near cathode region with a pure real and positive resistance which is independent from the frequency up to the plasma range, and the end of the dead zone coincides with the maximum value of velocity overshoot. The remaining part of the n -region, where SNDC takes place, can be considered as the active region of the diode. We remark that at low frequencies ($f = 55 \text{ GHz}$) $\text{Re}[\nabla Z(f)]$ is positive in each point of the diode (curve 1 in Fig. 11 (a)). By increasing the frequency, the active region with negative values of $\text{Re}[\nabla Z(f)]$ appears at first close to the anode and then widens and shifts towards the cathode. There, at sufficiently high frequencies several spatial regions with $\text{Re}[\nabla Z(f)] < 0$ can appear. In general, the maximum number of active regions which shows up in the spatial dependence of $\text{Re}[\nabla Z(f)]$ is equal to the number of generation bands in the frequency dependence of $\text{Re}[Z(f)]$. It is due to the fact that the curves in Fig. 11 correspond to the growing space-waves of the local electric field starting at the beginning of the active zone and vanishing at the anode contact.

To illustrate the usefulness of the impedance field for the designing of the diode, in the following we shall analyze the diode characteristics when the length of the n -region is reduced by keeping the same doping profile and total current. Figure 12 reports the effect of such a reduction on $\text{Re}[\nabla Z(f)]$ calculated at $f_3 = 290 \text{ GHz}$. Curve 1 corresponds to $l_n = 1.0 \mu\text{m}$ and curve 2 to $l_n = 0.9 \mu\text{m}$ when the anode n^+ -region is shifted to the left up to the first nearest point in which $\text{Re}[\nabla Z(f)]$ vanishes. We observe that, in doing so, the second generation band disappears, the new profile practically coincides with curve 1 in the common region, and only one active zones followed by a zone with positive values of $\text{Re}[\nabla Z(f)]$ remains. Moreover, $\text{Re}[Z(f)]$ of the whole diode becomes positive at $f_3 = 290 \text{ GHz}$ (see curve 2 in Fig. 10). To make $\text{Re}[Z(f)]$ at this

frequency negative again, it is necessary to shift the n^+ -anode contact to the second point where $\text{Re}[\nabla Z(f)]$ crosses the zero axis. This case is illustrated by curves 3 in Fig. 12 (analogously as in Figs 4 and 9) which is calculated for $l_n = 0.72 \mu\text{m}$. In this way one removes the near-anode region with $\text{Re}[\nabla Z(f)] > 0$ and, as a consequence, the diode can again generate at frequency $f_3 = 290 \text{ GHz}$ since its $\text{Re}[Z(f_3)]$ becomes again negative (see Fig. 10, curve 3). The generation band of the shorted diode is so extended to the higher frequency range which fully covers the second generation band of the initial diode.

This simple illustration proves that the diode can be considered as a sequence of connected in series zones which give separate but additive contributions to the net spectrum of the microwave power generation. Each contribution can be described by a local impedance-field. Under constant current-operation, the local characteristics of the diode (e.g., the drift velocity, the impedance field, etc.) are determined by the distance between the source (i.e. the cathode $n^+ - n$ boundary) and the local point of the structure and do not depend on the distance from the drain (i.e. the anode $n - n^+$ boundary). In other words, since the carrier flux starts at the source and ends at the drain, the local characteristics depend on the pre-history of carrier motion from the source only and contain no information about a further motion of carriers towards the drain. Because of that, the shortening of the n -region looks as a cut of the corresponding part of n -region which is on the anode side. (see Figs 4 and 12).

4. Conclusions

The deterministic nature inherent in the HD model is found to be quite appropriate for the calculation of the impedance field spectrum which describes the local profile of the small signal characteristics. The spatial dependence of the impedance field we have obtained in the whole frequency range of interest constructs a map which reflects the main physical processes occurring in the different regions of the device and can be used for several purposes such as: to give a comprehensive analysis of the device performance, to provide a proper choice of the device design, etc. Under current operation mode, the diode can be considered as a sequence of zones connected in series which give additive contributions to the amplification (and generation) spectrum. In the high-frequency region of generation, the main contribution comes from the zone which is near the anode. It is the zone where the transit-time dynamics of the accumulation layers takes place. The processes in the anode are found to have practically no influence on the diode performance. When the total current is kept constant, the reduction of the n -region length does not lead to any variation of the dead zone and implies primarily a shortening of the active zone only. This leads to a shift at higher frequencies of the generation spectrum. One can expect that a more significant variation of the generation spectrum can be obtained by a proper choice of the doping profile in this zone.

Acknowledgement

We would like to express our sincere gratitude to the Nordic Council of Ministers for financial support.

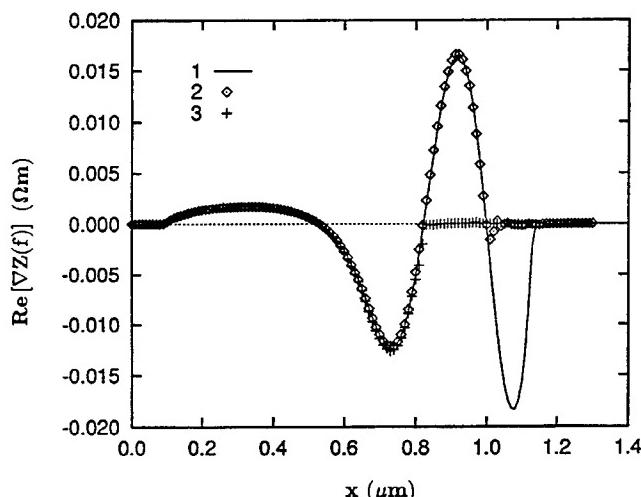


Fig. 12. Spatial profiles of the real part of the impedance field calculated at the frequency $f = 290 \text{ GHz}$. The notation is the same of Fig. 4.

References

1. di Forte-Poisson, M. A., *et al.*, Electron. Lett., **20**, 1061 (1984).
2. Rydberg, A., Kollberg, E., Electron. Lett., **17**, 770 (1986).
3. Rydberg, A., Electron Dev. Lett., **11**, 439 (1990).
4. Rydberg, A., Int. J. of Infrared and Millimeter Waves, **11**, 383 (1990).
5. Ekzhanov, A. E., Kosov, S. A. and Zotov, V. A., Proc. 20th European Microwave Conf. (Microwave Exhibition 1990), p. 1782.
6. Kamoua, R., Eisele, H. and Haddad, G. I., Solid-State Electron., **36**, 1547 (1993).
7. Tian, H., Kim, K. W., Littlejohn, M. A., Mishra, U. K. and Hashemi, M. J., Appl. Phys., **72**, 5695 (1992).
8. Gružinskis, V. *et al.*, Appl. Phys. Lett., **61**, 1456 (1992).
9. Mitin, V., Gružinskis, V., Starikov, E. and Shiktorov, P., J. Appl. Phys., **75**, 935 (1994).
10. Friscourt, M. R., Rolland, P. A., Cappy, A., Constant, E. and Salmer, G., IEEE Trans. Electron Dev., **ED-30**, 223 (1983).
11. Curov, M. and Hintz, A., IEEE Trans. Electron Dev., **ED-34**, 1983 (1987).
12. Wu, K. F., Czekaj, J. and Shaw, M. P., J. Appl. Phys., **74**, 315 (1993).
13. Gružinskis, V., Starikov, E. and Shiktorov, P., Solid-State Electron., **36**, 1055 (1993).
14. Gružinskis, V., Starikov, E. and Shiktorov, P., Solid-State Electron., **36**, 1067 (1993).
15. Gružinskis, V. *et al.*, Semicond. Sci. Technol., **8**, 1283 (1993).

Contacts to Monocrystalline N- and P-type Silicon by Wafer Bonding Using Cobalt Disilicide

Göran Thungström, Christer Fröjd, Per Svedberg* and C. Sture Petersson

Royal Institute of Technology, Dept. of Solid State Electronics; Electrum 229, S-164 40 Kista, Sweden

Received June 8, 1994; accepted June 14, 1994

Abstract

Contacts to monocrystalline silicon have been prepared by wafer bonding using cobalt disilicide as an interfacial layer. Bonding has been carried out with three different structures: $n^+ \text{-CoSi}_2 \text{-}n^+$, $p^+ \text{-CoSi}_2 \text{-}p^+$ and $p^+ \text{-CoSi}_2 \text{-}n^+$. The intermediate cobalt disilicide layers had a thickness of either 700 Å or 5250 Å. The bonding interface was characterized by electrical measurement (IV) and Secondary-ion mass spectrometry (SIMS) of the formed contacts. The $n^+ \text{-CoSi}_2 \text{-}n^+$ and $p^+ \text{-CoSi}_2 \text{-}p^+$ bondings display an ohmic behaviour. The resistance of the bonded structures was in the range expected for the bulk silicon used (0.1–0.05 Ω cm). The $p^+ \text{-CoSi}_2 \text{-}n^+$ structures shows a non ohmic behaviour. An evaluation of the SIMS profiles reveals that the non-linear behaviour of the $p^+ \text{-CoSi}_2$ -interface is due to phosphorous diffusion from the n-doped region across the silicide to the p-doped area. It is shown that the phosphorous compensates the boron dopant.

1. Introduction

The mechanism for cobalt disilicide wafer bonding is not fully understood. However, it is clear that the surfaces must be in close contact and that a reaction takes place. To insure contact the surfaces must be particle free and flat. An important factor for the bonding result is the attractive forces (e.g. van der Waals forces) between the surfaces that differ from one material to another. A strong attractive force can overcome the effects of rough surfaces. There are many surface related factors (material type, surface morphology, surface charges, adsorbed films etc.), which affect the adhesion forces. Glasses flow at relatively low temperatures which makes it easier for the surfaces to come within atomic distances. Reaction can then occur and result in strong bonding. The materials to be bonded must have similar thermal expansion coefficients in order to reduce stresses. The stresses can affect the performance and eventually break the bonding. The formation of a buried metallic-layer in silicon makes it feasible to produce buried interconnections and buried ground planes. In an earlier work we have investigated buried pattern Schottky contacts using cobalt disilicide as the interfacial layer [1]. Ohmic interconnections have been established with PtSi-PtSi bonding [2]. In this study we investigate the buried cobalt layer as an interconnection between different types of doped silicon. We have chosen CoSi₂ because of its low resistivity (18 μΩ cm), and good chemical resistance.

2. Experimental

When preparing the samples we work simultaneously with two wafers at time, one of them to be covered with cobalt. Six types of samples (Table I) were prepared in the following way. Silicon wafers with a diameter of 3", (100), 400 μm thick, n-type doped with phosphorus, and a resistivity of 0.1–0.05 Ω cm were prepared with a n^+ -layer by diffusion of phosphorus from a POCl_3 source. The diffusion was carried out at 1200 °C for 55 min. This treatment resulted in a highly doped phosphorus-layer with a concentration of approximately $8 \cdot 10^{20} \text{ cm}^{-3}$ and a depth of 3 μm. Silicon wafers with a diameter of 3" (100), 400 μm thick, p-type doped with boron, and a resistivity of 0.1–0.05 Ω cm were prepared with a p^+ -layer by diffusion of boron from a boron-nitride source. The diffusion conditions were 890 °C for 20 min in $\text{N}_2 + \text{O}_2$, for 2 min in $\text{N}_2 + \text{O}_2 + \text{H}_2$ and finally 60 min in N_2 . The drive-in step was 1200 °C for 25 min in $\text{N}_2 + \text{O}_2$, 105 min in $\text{N}_2 + \text{O}_2 + \text{H}_2$, and finally 30 min in $\text{N}_2 + \text{O}_2$. The resulting boron layer had a concentration of about $6 \cdot 10^{19} \text{ cm}^{-3}$ and a depth of 6 μm. The p-type wafer was dipped in buffered-HF for 40 min to remove the oxide. After that the wafer was dipped in 1:10 HF-solution and dried in blowing N_2 . The n-type wafer was dipped in 1:10 NF-solution and dried in N_2 . The wafer to be covered with cobalt (see Table I) was immediately loaded into a load-lock chamber and finally into the vacuum-chamber with a base-pressure of about $1 \cdot 10^{-8} \text{ mbar}$. The cobalt layer was deposited via e-beam evaporation. After the evaporation the wafer was kept in the load-lock chamber until the second wafer was ready for the bonding process. The wafers were pressed together and a quartz-plate weighing 100 gram was placed on top of the wafers in order to counteract the bending caused by the difference in the thermal expansions of the buried layer and the silicon

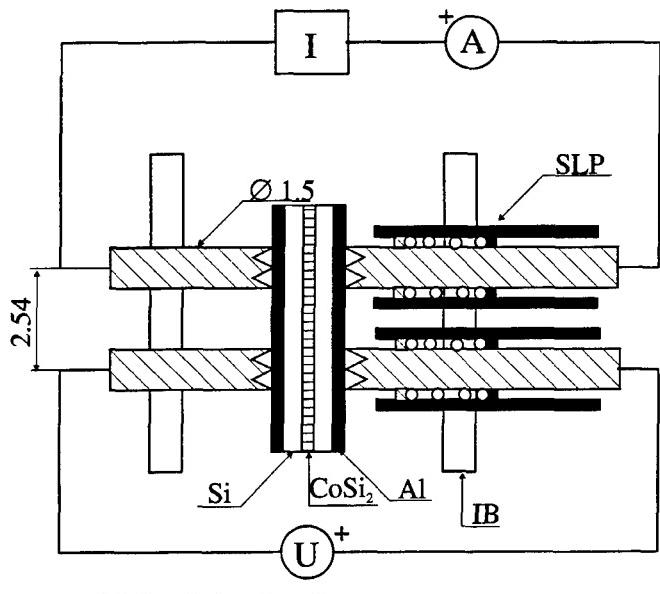
Table I. Preparation of the samples

Code	Structures	Thickness of the cobalt-layer (Å)	Heat-treatment 900 °C
A	$n^+ \text{-Co-}n^+$	1500	1 h
B	$p^+ \text{-Co-}p^+$	1500	1 h
C	$n^+ \text{-Co-}n^+$	200	30 min
D	$p^+ \text{-Co-}p^+$	200	30 min
E	$p^+ \text{-Co-}n^+$	1500*	1 h
F	$p^+ \text{-Co-}n^+$	200*	30 min

* Evaporated on the p-type wafer

* Present address: Per Halvledare, Hässelbyvägen 20, S-163 52 Spånga, Sweden.

e-mail address: goran@ele.kth.se



IB: Insulating board

SLP: Spring loaded probe

Fig. 1. The four point probe and electrical characterization of n^+ - CoSi_2 - n^+ and p^+ - CoSi_2 - p^+ samples.

wafers. The bonding/silicidation was carried out in N_2 at 900°C in a furnace. The bonded samples were then dipped in a 1:10 HF-solution to remove the thin oxide grown on the back sides. An Al-layer, 1 µm thick, was deposited on both sides, and annealed at 515°C for 20 min in N_2 to create good ohmic contacts. The bonded wafers were then cut into pieces with a size about 16 mm².

3. Characterization

3.1. IV-measurement

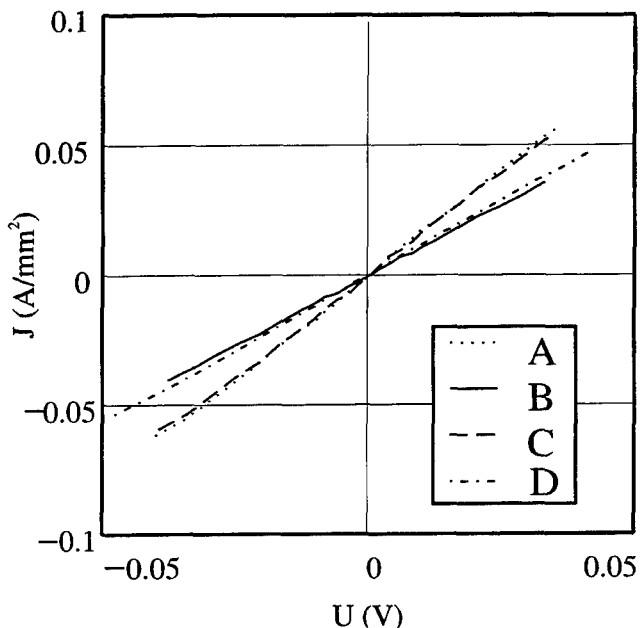
The contacts were electrically characterized with a specially designed four point probe. Figure 1 shows how to probe the sample and the main features of the four point probe. The current source has a maximum current limit of 1.25 A. The current and voltage are measured using a Fluke 8060A as a current meter and a Hewlett Packard 34401A as a voltage meter. When the voltage and current probes are switched no measurable changes in the voltage were observed. The area of the sample was measured using a calliper with a resolution of 0.05 mm. The measurements were carried out at a temperature of 295K.

3.2. SIMS-measurement

The characterization of the dopant concentration was made by using an secondary-ion mass spectrometer (Cameca 4f) O_2^+ as the primary ion and detecting $^{11}B^+$, $^{31}P^+$ and $^{59}Co^{2+}$.

4. Results

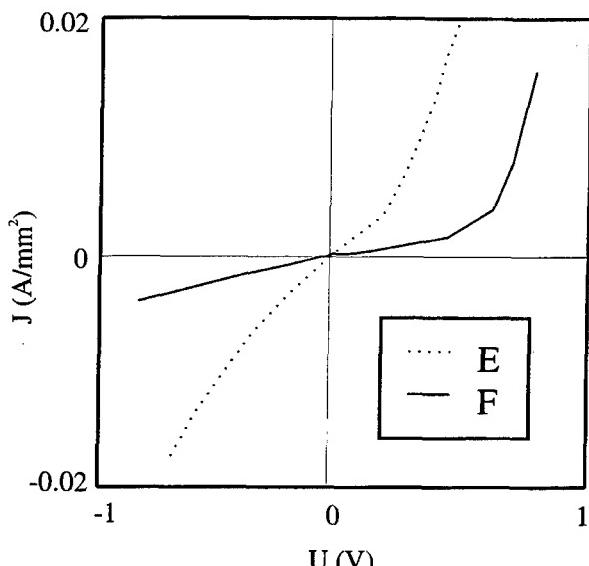
Figure 2 shows the current density vs. voltage for the samples n^+ - CoSi_2 - n^+ and p^+ - CoSi_2 - p^+ . The graphs are

Fig. 2. The current density vs. voltage for the n^+ - CoSi_2 - n^+ and p^+ - CoSi_2 - p^+ samples.

linear and show different slopes for the n^+ - CoSi_2 - n^+ and p^+ - CoSi_2 - p^+ samples. The slopes are estimated by fitting straight lines with the least square method. The data are summarised in Table II. The p^+ - CoSi_2 - n^+ samples show a

Table II. The measured data of n^+ - CoSi_2 - n^+ and p^+ - CoSi_2 - p^+ samples

Code	Bonded structures	Silicide-thickness (Å)	Sample area (mm ²)	Slope (Ω mm ²)	Calculated resistivity (Ω cm)
A	n^+ - CoSi_2 - n^+	5250	14.4	0.63	0.051
B	p^+ - CoSi_2 - p^+	5250	22.5	0.93	0.075
C	n^+ - CoSi_2 - n^+	700	15.0	0.65	0.052
D	p^+ - CoSi_2 - p^+	700	16.8	0.88	0.071

Fig. 3. The current density vs. voltage for the p^+ - CoSi_2 - n^+ samples.

non-ohmic behaviour as can be seen in Fig. 3. A further investigation of the p^+ -CoSi₂-n⁺ sample with a thicker silicide (5250 Å) was carried out to characterize the non-ohmic behaviour. A small part of the sample was etched out from one side to remove the silicon down to the silicide-layer. In Fig. 4 a schematic drawing demonstrates the electrical probing of the sample. It is the same set-up as earlier even if the "four point" geometry was not used. Figure 5 shows how to make an estimation of the resistance between the probes and the back side Al contacts. The contact resistance can then be calculated as $2R_c = (U_y - U_i)/I$. The voltage over the p^+ -CoSi₂ contacts can be estimated as $U_p = U_p' - R_c I$, and voltage over n⁺-CoSi₂ contacts $U_n = U - U_p' - R_c I$. Figure 6 shows that the non-ohmic behaviour is caused by the p^+ -CoSi₂ interface. For SIMS analysis,

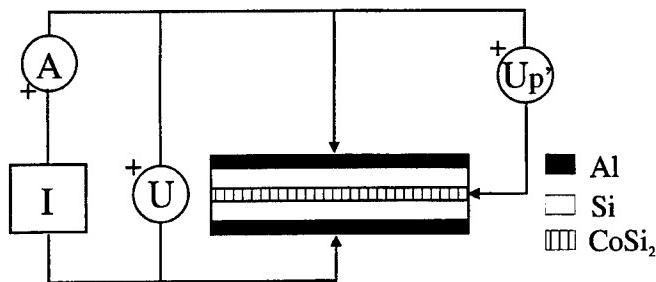


Fig. 4. Electrical characterization of the p^+ -CoSi₂-n⁺- and CoSi₂-n⁺ interfaces for the p^+ -CoSi₂-n⁺ sample with 5250 Å CoSi₂-layer.

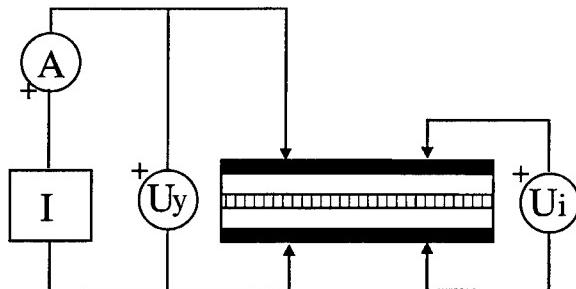


Fig. 5. Sketch illustrating the method used to measure the voltage drop between the Al back side contacts and the probes.

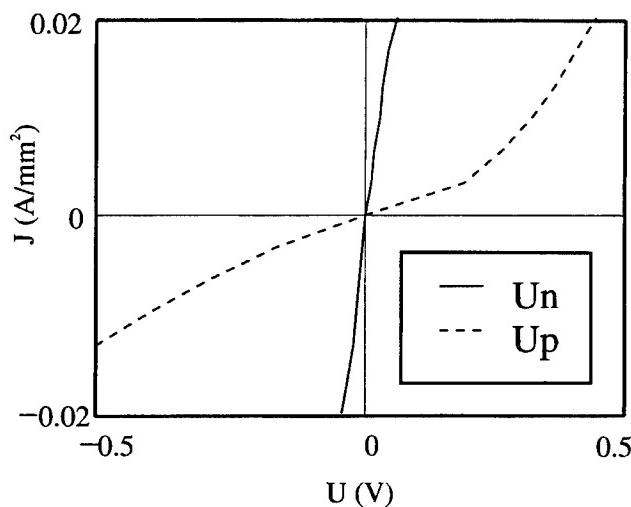


Fig. 6. The current density vs. voltage for the p^+ -CoSi₂-n⁺ and CoSi₂-n⁺ interfaces.

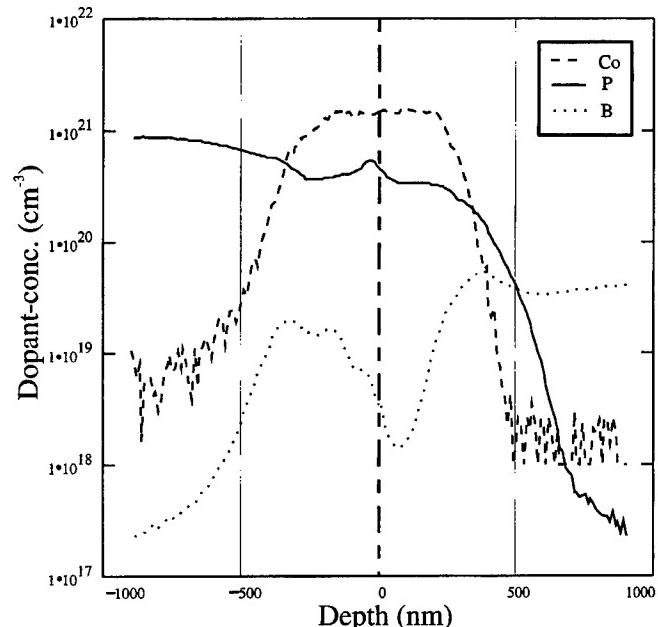


Fig. 7. The dopant concentration vs. depth characterized with SIMS.

one sample was etched down to the silicide through the n-layer and a second sample was etched down to the silicide through the p-layer. Figure 7 shows (sample E) the boron and phosphorus profiles in the p^+ -CoSi₂-n⁺ layer. The cobalt signal indicates the position of the CoSi₂-layer. Phosphorus has diffused over the p-side and by compensating the boron has created a p-n junction.

5. Discussion

The n⁺-CoSi₂n⁺ and p⁺-CoSi₂-p⁺ samples display an ohmic behaviour. The resistances of the sample consist of several parts: the contact resistance in the Al-silicon interfaces, the bulk resistance in the silicon, the contact resistance in the silicide-silicon interfaces and finally the resistance of the silicide layer. It is not possible to divide the resistance into their separate components using only one measurement. Because the estimated resistance of the bonded samples correspond to that calculated from the resistivity of the wafers (0.1–0.05 Ω cm), the contact resistance cannot be estimated. The different slopes of n⁺-CoSi₂-n⁺ and p⁺-CoSi₂-p⁺ samples are most likely the result of the different bulk resistivities of the n-type and p-type wafers. The p⁺-CoSi₂-n⁺ silicide sample had a non-ohmic behaviour originating in the p^+ -CoSi₂ interface. As in n⁺-CoSi₂-n⁺ samples, the n⁺-CoSi₂ interface displays an ohmic behaviour. The SIMS measurements show that the non-ohmic behaviour is caused by diffusion of the phosphorus through the bonding interface. The phosphorus concentration is significantly higher than that of boron in the p^+ -CoSi₂ interface. The boron profile shows a drop in the CoSi₂, which is not well understood. Boron diffuses easily at 900 °C in cobalt disilicide, therefore a uniform concentration profile would be expected [3]. Similarly boron profiles in cobalt disilicide have been previously reported [4] similarly unexplained. Should one desire an ohmic behaviour for p^+ -CoSi₂-n⁺, this could then be achieved by increasing the

boron concentration and decreasing the phosphorus concentration. The phosphorus concentration must be estimated with respect to the tunneling range and the expected contact resistance of the silicide-silicon interface. Another way to solve the problem could be to use a diffusion barrier or change to another n-dopant possibly antimony, which does not diffuse easily in CoSi₂ [3].

6. Conclusions

We have shown that it is possible to create ohmic contacts with n⁺-CoSi₂-n⁺ and p⁺-CoSi₂-p⁺ structures using wafer bonding. The contact resistance when compared to the bulk silicon resistance is negligible. Contacts with p⁺-CoSi₂-n⁺ structures show a non-ohmic behaviour. Electrical measurement of the p⁺-CoSi₂ and n⁺-CoSi₂ interfaces indicate clearly that the non-ohmic behaviour is caused by p⁺-CoSi₂.

interface. The SIMS measurement shows that this non ohmic behaviour is caused by diffusion of the phosphorus through the bonding/silicide interface.

Acknowledgement

The authors gratefully acknowledge Dr. F. M. d'Heurle for valuable discussion. We thank M. Linnarsson for running the SIMS analysis.

References

1. Thungström, G., Fröjd, C. and Petersson, C. S., To be published in Mat. Res. Soc. Symp. Proc. **337** (1994).
2. Ismail, M. S. and Bower, R. B., Electronic Lett. **27**, 1153 (1991).
3. Thomas, O. et al., J. Appl. Phys. **64**, 2973 (1988).
4. Zaring, C., Svensson, B. G. and Östling, M., Mat. Res. Soc. Symp. Proc. **260**, 157 (1992).

Magneto Luminescence of As-Grown InAs/InP Quantum Well Islands

H. Sigg

Paul Scherrer Institut Zürich, Badenerstr 569, CH-8048 Zürich, Switzerland

P. Christianen

High Field Magnet Laboratory and Research Institute for Materials, University of Nijmegen, 6525 ED Nijmegen, The Netherlands

and

R. Houdré and A. Rudra

Institute for Micro and Optoelectronics, EPF Lausanne, CH-1010 Lausanne, Switzerland

Received June 11, 1994; accepted June 15, 1994

Abstract

InAs/InP quantum well structures with an epitaxially controlled amount of island growth are investigated by excitation spectroscopy and magneto-luminescence. From the diamagnetic shift of the luminescence, an approximate value of the exciton radius is obtained. The very small lateral radius of below 6 nm is evidence of reduced dimensionality and lateral confinement effects.

The subject of self-organized growth in the epitaxy of group III-V material systems has recently become an exceptional challenge to the crystal growing community. Having tried in the past first to grow highly homogeneous quantum wells (QW) structures and later to laterally process them into quantum wires or quantum dots, the community is now searching for growing procedures yielding the same result in a single step. This search requires specialized in- and ex-situ characterization methods to provide a detailed knowledge of the structural and the electronic properties. Here we analyze the method of magneto-optical spectroscopy for elucidating the lateral size quantisation and structural properties of strained InAs/InP QW systems. Measurement in magnetic fields up to 20 T are performed.

We investigate for the first time systems with an *a priori* unknown confinement factor obtained by self-organized island growth [1]. The island size could so far not be determined from structural analyses, but it may fall in the range of several μm down to fractions of a nm. It has been recently proposed that the diameter of the island in these systems is of the order of 10 nm, sufficient for lateral confinement [2].

The ultra-thin InAs/InP QW are grown by chemical beam epitaxy (CBE). On top of a $0.25\ \mu\text{m}$ thick InP buffer layer, a single, two monolayer thick InAs QW layer is deposited. It is found that a growth interruption of approximately 10 s leads to a spontaneous deformation of the layer into a manifold of QW islands of different thicknesses [1, 3]. Samples produced without growth interruption show the typical 2D reflection high energy electron diffraction (RHEED) pattern. The sample of this kind investigated here will be refer to as sample B. The RHEED pattern changes during the growth interruption into a spotty picture along

the (011) azimuth, attributed to islands elongated along (011).

A typical low-temperature photoluminescence spectrum of a growth-interrupted sample (sample A) is shown in Fig. 1. The multiple peaks correspond to emission of QW's of different, but well-defined thicknesses. Applying a simple bandstructure calculation taking into account strain effects [4], the following assignment can be made [1, 2]: the luminescence peak at 1020 nm corresponds to emission from two monolayer thick QW's, while the emission peaked at lower energies corresponds to 3, 4, ... etc. monolayer QW's. The so-called excitation spectrum, Fig. 1 to the left, shows efficient excitation at energies beyond the bandgap of InP. The detection wavelength is set to 1280 nm of the 5 mono layer thick QW. The efficiency continuously drops off for energies below the InP gap. Because of the weak absorption strength and because of the hampered carrier transfer between disconnected islands, no excitation into any of the thinner wells is observed. Information about the island diameter could thus not be obtained from the excited level spectroscopy.

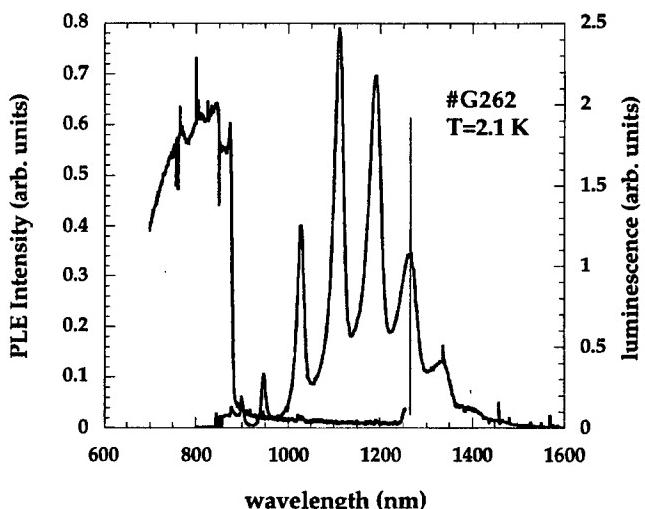


Fig. 1. Luminescence and photo luminescence excitation intensities of a two monolayer thick InAs/InP quantum well grown with growth interruption, sample A.

Instead, the magneto-luminescence technique is applied, where use is made of the fact that the diamagnetic shift of an exciton depends on its reduced effective mass μ , and is proportional to the exciton diameter in the plane *normal* to the applied magnetic field [5–8].

Measurements were thus taken with the magnetic field oriented in 3 major crystallographic directions defined as follows: the z-direction is normal to the QW layer (001), and x, y are the two perpendiculars in plane directions parallel to the cleaved edges. The HF-induced etch pit V-grooves on the *substrate* side are directed along the x-axes defining this axis as the (011).

Measurements were performed at low temperature, $T \approx 5\text{ K}$ and at magnetic fields of up to 20 T. Low-field spectra (up to 11 T) were recorded by a Bruker 113 v Fourier transform spectrometer, where the sample luminescence was imaged from the spectrometer exit port onto an InGaAs p-i-n detector located in front of the interferometer chamber. The high-field spectra were obtained from scanning a HR640 single monochromator equipped with a cooled Ge-detector.

The peak position of the luminescence was determined by fitting the luminescence intensities in the vicinity of the peak to a function consisting of a Lorentzian superimposed on a small linear background. The line shape, intensities and background level were found to be independent of the magnetic field. The magnetic field dependence of the luminescence can thus be described by a single free parameter: the peak position E . In all three directions we observe that the peak positions shift as $E_i = E_0 + m_i B^2$, where m_i is the anisotropic diamagnetic shift parameter. This is shown for the 4 mono-layer QW in Fig. 2 neglecting $E^{(0)}$ for clarity reasons. It should be noted that we had to reposition the samples in order to achieve the three experimental arrangements. This caused an uncertainty in $E^{(0)}$ of approximately 3 meV.

A first indication for localization in our InAs/InP structures is obtained from the comparison of the results shown in Fig. 2 with diamagnetic shift parameters as measured in

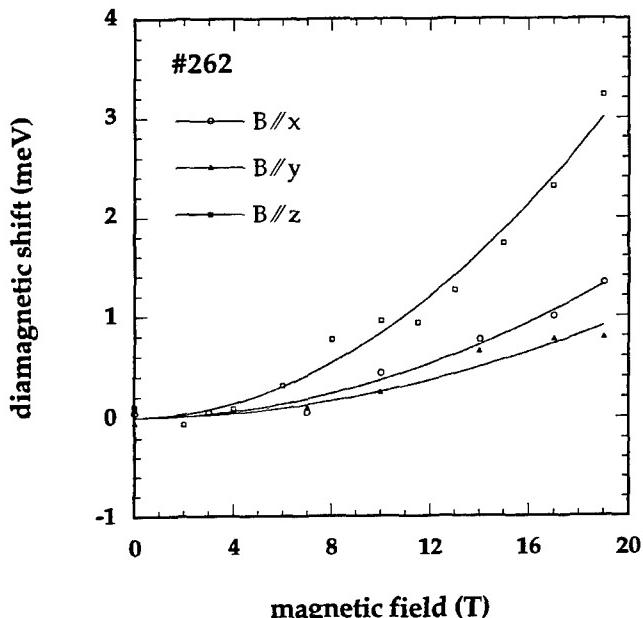


Fig. 2. Diamagnetic shift of the emission line peaked at approximately 1.052 eV of sample A.

crescent-shaped GaAs/AlGaAs quantum wires of known confinement. Our values for respectively m_x , m_y , and m_z of approximately 3.2, 2.9 and $8.3\text{ }\mu\text{eV/T}^2$ are much smaller than the values measured for the QW wires of 7, 4.5 and $16.3\text{ }\mu\text{eV/T}^2$, for wire dimensions of approximately 20 nm by 10 nm [5], and also much smaller than the values of 12, 4 or less and $20\text{ }\mu\text{eV/T}^2$ for wires of approximately 35 nm by 5.5 nm [6]. Our second qualitative argument is based on the observed anisotropy. The diamagnetic shift along the x direction is slightly larger than that along the y direction, suggesting an in-plane cigar-shaped exciton, elongated along the same direction as that determined by RHEED. We attribute this shape to an anisotropic confinement, since we expect that such embedded structures are isotropically stressed. The anisotropy is, however, rather weak and is observed only in the measurements at high magnetic fields. Furthermore, no polarization of the emitted light was observed.

It has been argued [5] that under given circumstances, the diamagnetic shift of a laterally confined exciton can be described analog to the exciton in a biaxial crystal [9]. The anisotropic confinement and resulting band-mixing [10], which normally complicate the evaluation are phenomenologically introduced in the form of an anisotropic exciton mass μ_i ($i = x, y, z$). The exciton radius a_i is estimated as usual from $a_i = (4\pi\hbar^2\varepsilon/e^2\mu_i)$, where ε is the dielectric constant, and where the other constants have their usual meaning. The reduced masses are obtained from the magnetic field dependence of the peak position given by $m_i = (kB^2/\mu\mu_j\mu_k)$, where μ is an appropriate average over all directions $\mu^{-1} = 3^{-1}(\mu_x^{-1} + \mu_y^{-1} + \mu_z^{-1})$, and k is given by $4\pi^2\hbar^4\varepsilon^2/e^2$. The calculated average exciton mass and the exciton radius $a_{x,y}$ defined as $\sqrt{a_x a_y}$, are plotted in Fig. 3.

The direction of smallest extension is, as expected, along the growth direction. We also obtain that the exciton radius a_z increases with decreasing well width. We interpret this as due to leakage of the electron wave function into the InP barrier. The average in-plane radius is almost independent of the energy, i.e. the QW thickness, and its typical value of about 5.5 nm compares very well with the radii obtained for QW wires [5] of 8.6 nm along and 5.5 perpendicular to the wires. There the radius a_z was determined to be 2.4 nm,

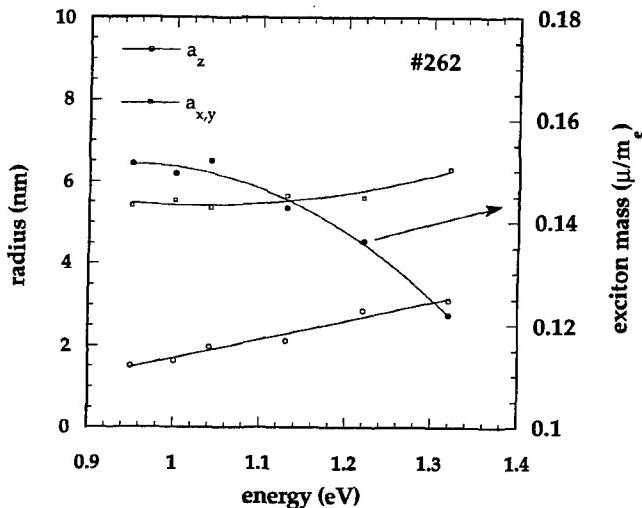


Fig. 3. Exciton radius parallel (a_z) and perpendicular ($a_{x,y}$) to the growth direction and averaged exciton mass determined from the diamagnetic shift of the luminescence of sample A.

Table I. Measured diamagnetic shift parameters and calculated anisotropic exciton radii

sample	position (eV)	m_x (μeV/T ²)	m_y (μeV/T ²)	m_z (μeV/T ²)	a_x (nm)	a_y (nm)	a_z (nm)
A	1.12	4.2	3.2	9.8	4.9	6.5	2.1
B	1.11	3.2	3.5	8.3	5.5	5.1	2.5

which is the same value as we obtain for the 3 monolayer QW case.

Striking, however, is the observed similarity of the results from samples A and B, shown in Table I. This agreement suggests that the homogeneity on the length scale of an exciton does not depend on the growth interruption, and thus on island formation. From the comparison with the better defined quantum wire systems, we are led to believe that the excitonic recombination is governed by confinement effects. From this we conclude that the InAs-layer of sample B is also not homogeneous. Indeed the recombination consists of a doublet. For the slightly stronger high-energy peak at 1.16 eV, we obtain an in-plane radius of approximately 7 nm, the highest value of all the investigated features.

Unexplained at this stage is the fact that the RHEED analyses is suggestive of a 2D-growth mechanism for sample B. This technique is perhaps only sensitive to larger scale features and does not resolve short range fluctuations, which however could cause the observed confinement.

We conclude that the diamagnetic shift of the magneto-luminescence in InAs/InP QW, although tiny, can be resolved by measurements performed in high magnetic fields. The obtained exciton radii are close to those observed in quantum wires of well-defined size. The oversimplified

model of an anisotropic effective mass exciton is found to form a useful basis for discussions. The origin of the confinement, since it is observed in samples grown with and without growth interruption is still not understood.

References

1. Carlin, J. F., Houdré, R., Rudra, A. and Illegems, M., *Appl. Phys. Lett.* **59**, 3018 (1991).
2. Brasil, M. J. S. P., Nahory, R. E., Tamargo, M. C. and Schwarz, S. A., *Appl. Phys. Lett.* **63**, 2688 (1993).
3. Houdré, R., Carlin, J. F., Rudra, A., Ling, J. and Illegems, M., *Superlattices and Microstructures* **13**, 67 (1993).
4. Marzin, J. Y., in "Heterojunctions and Semiconductor Superlattices" (Edited by G. Allen, G. Bastard, M. Lannoo and M. Voos) (Springer, Berlin 1986), p 161.
5. Nagamune, Y. *et al.*, *Phys. Rev. Lett.* **69**, 2963 (1992).
6. Plaut, A. S., Kash, K., Kapon, E., Hwang, D. H. and Colas, E., *Proc. EP2DS-10* (1993), *Surface Science* **305**, 576 (1994).
7. Sakaki, H., Arakawa, Y., Nishioka, M. and Yoshino, J., *Appl. Phys. Lett.* **46**, 83 (1985).
8. Weman, H. *et al.*, *Superlattices and Microstructures* **13**, 5 (1993).
9. Taguchi, S., Goto, T., Takeda, M. and Kido, G., *J. of Phys. Soc. Jpn.* **57**, 3256 (1988).
10. Bauer, G. E. W. and Ando, T., *Phys. Rev. B* **38**, 6015 (1988).

Initial Strain Relaxation and Optical Quality in Lattice Mismatched InGaAs/GaAs Single Quantum Wells

S. M. Wang and T. G. Andersson

Department of Physics, Chalmers University of Technology, S-41296 Göteborg, Sweden

Received May 2, 1994; accepted June 15, 1994

Abstract

We have studied initial strain relaxation and optical quality in lattice mismatched $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ single quantum wells as a function of In content, x , using several structural and optical characterisation techniques. Influences of cap layer thickness on material qualities were also examined. Samples were grown with solid source molecular beam epitaxy. The initial relaxation was predominated by formation of misfit dislocations for $x < 0.25$ and three-dimensional (3D) islands for $x \geq 0.3$. Both phenomena were observed in the transition region. Presence of misfit dislocations gave rise to degraded optical qualities in terms of luminescence efficiency and linewidth. In the case of 3D islanding, the optical signal was strong but broadened or even split into several peaks. Critical layer thickness (CLT), defined as onset of strain relaxation decreased with In content. The x -dependent CLT was explained by energy consideration. Capping a GaAs layer on a partially relaxed alloy overlayer increased the residual strain and thus improved optical qualities.

It is well known that lattice misfit in semiconductor heterostructures, e.g. a thin InGaAs overlayer grown on a GaAs substrate, can be elastically accommodated when the alloy layer thickness is less than a critical value [1]. After the critical layer thickness (CLT) is reached, the structure relaxes and its material qualities deteriorate. There have been a number of experimental evidences showing that the initial strain relaxation in this system is caused by formation of misfit dislocations at the heterointerface when the In content, x , is less than 0.25 [2–6], and by a morphological change from two-dimensional layer-by-layer growth to three-dimensional (3D) island nucleation for larger x -values [7–14]. Use of surface sensitive techniques, such as scanning tunnelling microscopy (STM), reflection high energy electron diffraction (RHEED) and transmission electron microscopy (TEM), has revealed that these islands can be either dislocation free [14] or nucleated on misfit dislocations [10]. Thus the growth mode of $\text{In}_x\text{Ga}_{1-x}\text{As}$ follows the Stranski-Krastanov mode [15]. A quantum well (QW) structure is formed when the alloy is capped by a GaAs layer. Understanding strain relaxation in such embedded structures are of practical importance, since they are employed in device structures. The GaAs cap layer not only exerts a strain force on the alloy layer same as the GaAs substrate, but also induces In segregation [16]. Hence, both interface morphology and strain relaxation in QW structures are different from those in thin alloy overlayers.

In this paper we report initial strain relaxation and optical quality in lattice mismatched $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ single QWs for the whole In content range. We have used TEM and photoluminescence (PL) to examine structural and optical qualities, since STM and RHEED are unable to detect features on the interfaces and in the alloy layer beneath the sample surface. Other techniques, such as

double crystal X-ray diffraction, secondary ion mass spectroscopy (SIMS) and PL excitation etc., have also been used. Since the spatial resolution of PL is limited by the exciton Bohr diameter, typically about 300 Å, the PL is less sensitive when there are small islands as compared with TEM and RHEED. The PL resolution of measuring strain relief by formation of misfit dislocations is determined from the excitonic peak shift. In an ideal sample with uniform alloy compositions and strain relaxation, a peak shift of 1 meV corresponds to strain relief of about $(\Delta a/a)_\parallel = 2 \cdot 10^{-4}$. Critical layer thickness, defined as the onset of either dislocation formation or 3D islanding, was measured as a function of In content within an experimental resolution. The results were explained using energy considerations. We also found that the GaAs cap layer thickness had significant influence on strain recovery and optical quality.

A large number of samples were grown by molecular beam epitaxy in both Varian GEN II and 360 systems. Calibrations of growth rates and alloy compositions were made by RHEED oscillations. In each sample a 0.5 μm thick GaAs buffer layer was first deposited at 580 °C. The growth temperature was then reduced to 520–530 °C for the remaining parts. Both a pyrometer and a specially designed probe [17] were used to measure growth temperature. Detailed experimental set-ups for different diagnostic techniques have been reported before [18–20].

Misfit dislocations were directly observed in plan-view TEM images at a initial stage of relaxation for $x < 0.3$. Figure 1(a) shows a typical TEM picture image under the two beam condition of $g(220)$ for a partially relaxed, 250 Å thick $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}/\text{GaAs}$ single QW capped with a 50 Å thick GaAs layer. Misfit dislocations appear as an array and distribute asymmetrically along the two $[\pm 110]$ directions. Using the $\mathbf{g} \cdot \mathbf{b} \neq 0$ rule, where \mathbf{b} is the Burgers vector of misfit dislocations, we identified that they were of 60° type. The asymmetric distribution was also reported before [4, 21] and was attributed to different nucleation rates and gliding velocities of α and β dislocations [22]. The average distance between adjacent dislocations is about $0.8 \pm 0.3 \mu\text{m}$, resulting in a strain relaxation degree of $(2.5 \pm 1) \cdot 10^{-4}$. Figure 1(b) shows PL spectra from the same sample and the reference sample, which has a 5000 Å thick cap layer and is nearly fully strained. For the relaxed structure, the peak energy shifts to a lower energy, the linewidth broadens and the intensity decreases. These spectral features indicate degraded optical quality. The red shift of the PL transition energy is indicative of a reduced alloy band gap. Its magnitude is nearly proportional to the residual strain.

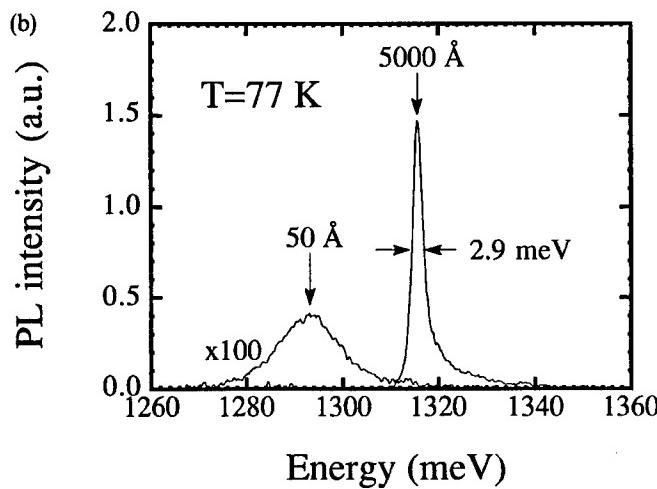
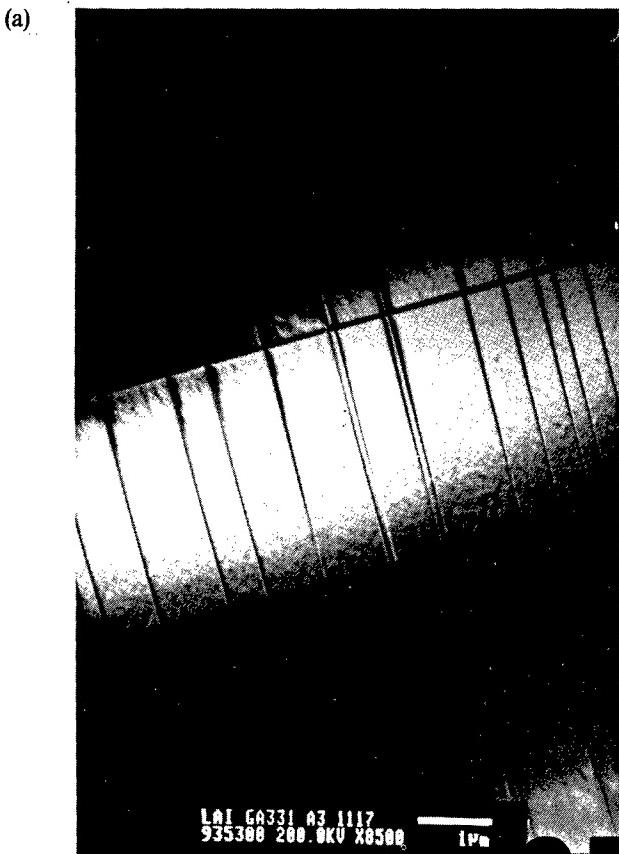


Fig. 1. (a) Plan-view TEM image of a 250 Å thick $\text{In}_{0.2}\text{Ga}_{0.8}\text{As}/\text{GaAs}$ single QW capped by a 50 Å thick GaAs layer. Misfit dislocations are clearly observed as an array and distributed asymmetrically along the two diagonal directions. (b) Photoluminescence spectra from the same sample shown in (a) and from the reference sample with a 5000 Å thick cap layer, which has a strong and narrow PL-peak.

Misfit dislocations trap photo-excited carriers and therefore the radiative recombination is reduced. Moreover, dislocation formation causes a non-uniform strain relaxation and thus broadens the luminescence peak.

For $x \geq 0.3$, 3D islands were observed to prevail at initial strain relaxation. Figure 2(a) shows a cross-section TEM image from $\text{In}_{0.36}\text{Ga}_{0.64}\text{As}/\text{GaAs}$ single QWs with 10, 20 and 30 Å thick alloy layers separated by 200 Å thick GaAs. For the 10 and 20 Å thick QWs, both upper and lower interfaces are flat. The corresponding PL signals (not shown) have a strong intensity and a narrow linewidth. The 30 Å

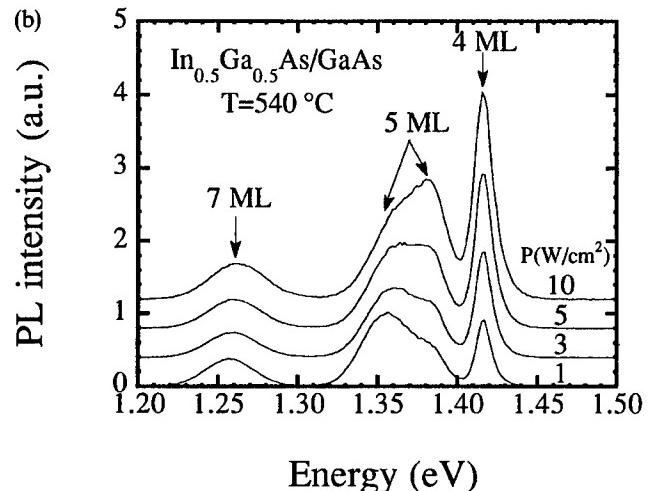
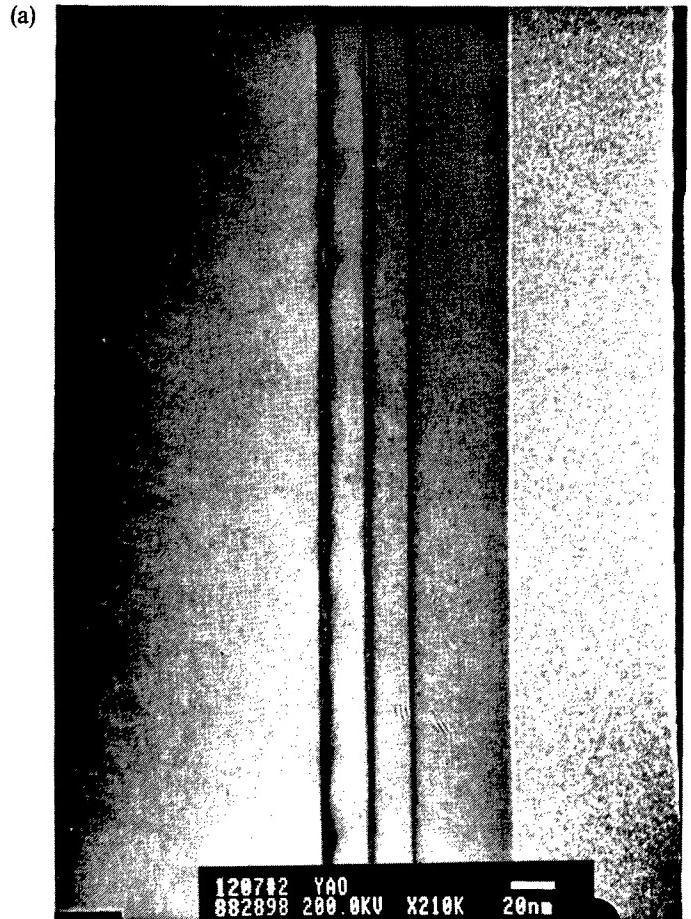


Fig. 2. (a) Cross-section TEM image of $\text{In}_{0.36}\text{Ga}_{0.64}\text{As}/\text{GaAs}$ single QWs with well thicknesses of 10, 20 and 30 Å. A wavy interface observed in the 30 Å thick QW indicates the presence of 3D islands. (b) Photoluminescence spectra from $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}/\text{GaAs}$ single QWs with 4, 5 and 7 monolayer thick alloy layers under different excitation intensities. The 3D islanding starts in the 5 monolayer QW. The related PL shows a resolved peak split.

thick QW, however, shows a flat lower interface, while the upper one is wavy, indicating presence of 3D islands. The average size of these islands is larger than 200 Å, i.e. comparable to the exciton Bohr diameter. Formation of 3D islands could cause strain relaxation [14]. This is non-uniform, large on top and in the middle of an island and small at its peripheral. Typical PL spectra in the presence of 3D islands are shown in Fig. 2(b) from $\text{In}_{0.5}\text{Ga}_{0.5}\text{As}/\text{GaAs}$ single QWs under different excitation intensities. Three-dimensional

islanding occurs in the 5 monolayer thick QW. The PL signal is broad and two peaks are resolved. The lower energy peak dominates at low excitations, while it saturates and the higher energy peak emerges at high excitations. Since the transition energy in the island areas is smaller due to a relatively larger well thickness, the excitons are easily trapped in such areas to form localized excitons. Increasing excitation intensity will saturate this localized exciton recombination and the carriers start to occupy states in areas between islands. The related excitonic transition has a higher energy because of a thinner well thickness. The PL signal then shifts to a higher energy. In general, presence of 3D islands does not significantly reduce luminescence efficiency. The peak split is resolved when the average lateral size of 3D islands is much larger than the exciton Bohr diameter.

Using the PL and TEM techniques, we have examined the CLT in $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ single QWs for the whole In content range. The results are summarized in Fig. 3. The CLT is the virtual boundary between the solid dots denoting dislocation-free two-dimensional alloy layer and the open circles (3D islands) or crosses (misfit dislocations). The initial strain relaxation is caused predominantly by formation of misfit dislocations for $x < 0.25$ and by 3D islands for $x \geq 0.3$. There is a transition region for $0.25 \leq x < 0.3$, in which both features are concomitant. Strain relief by formation of misfit dislocations was predicted a long time ago [23]. Matthews and Blakeslee suggested [24] that misfit dislocations were formed at interfaces as kinks. Assuming that the misfit between GaAs and InGaAs, f , is accommodated by both elastic strain, ϵ , and misfit dislocations, minimizing the sum of strain and dislocation energy and setting $\epsilon = f$ yield an expression for CLT, L_{dis} , as

$$L_{\text{dis}} = \frac{b}{4\pi f} \frac{1 - \frac{v}{4}}{1 + v} \ln \left(\frac{\alpha L_{\text{dis}}}{b} \right) \quad (1)$$

for double-kink misfit dislocations, where b is edge component of the Burgers vector, v is the Poisson ratio and α is a dislocation core cut-off parameter which is approximately 4. This is shown by the solid line, L_{dis} , in Fig. 3 which fits the CLT for $x < 0.3$.

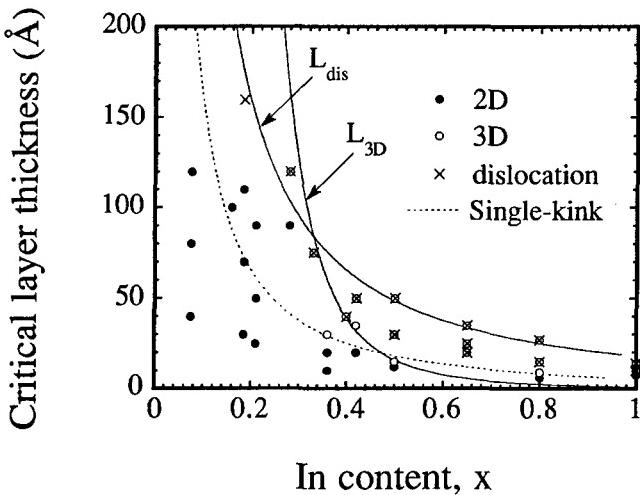


Fig. 3. Summary of strain relaxation and CLT in $\text{In}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ single QWs as a function of In content. Symbols are experimental data while the lines are the theoretical predictions.

A similar approach is applicable to the case of 3D islands. Decrease of strain energy by forming 3D islands occurs at the expense of increased surface energy. Since the strain relief within an island is non-uniform and depends on its geometrical parameters, it is not straightforward to establish a relation between the strain energy and the characteristic geometrical sizes to make a quantitative comparison. An alternative choice is to use the energy balance. The CLT, L_{3D} , is defined when the amount of reduced strain energy and increased surface energy are equal. The residual strain at the CLT then drops abruptly from the f -value, depending on the strain relaxation efficiency. This consideration is simple and intuitive, though the obtained L_{3D} -value could be larger than that from the energy minimization. The reduced strain energy per area, ΔE_e can be written as

$$\Delta E_e = 2\eta G f^2 L_{\text{3D}} \frac{1 + v}{1 - v} \quad (2)$$

where G is the shear modulus and η is the relaxation efficiency of 3D islands. Modelling the islands as hemispheres with radius, r , the increased surface energy per area, ΔE_s , is

$$\Delta E_s = d\pi r^2 \gamma \quad (3)$$

where d is the sheet density of 3D islands and γ is the surface tension. The island radius is related to film thickness and island density by

$$r = \left(\frac{3L_{\text{3D}}}{2\pi d} \right)^{1/3} \quad (4)$$

Combining eqs (2)–(4), the CLT defined at the onset of 3D island growth follows

$$L_{\text{3D}} = \frac{9\pi d}{32\eta^3} \left(\frac{\gamma}{G} \right)^3 \left(\frac{1 - v}{1 + v} \right)^3 \frac{1}{f^6} \quad (5)$$

For simplicity, we assume $\eta = 1$, i.e. the 3D islands are fully relaxed. The island density has been measured by STM [14]. The results revealed that the density was approximately proportional to x^2 . After substitution of numerical values in eq. (5), the CLT has a simple dependence on x , in Å

$$L_{\text{3D}} \approx \frac{1}{x^4} \quad (6)$$

This curve is denoted by L_{3D} in Fig. 3. The CLT decreases with In content and is below the L_{dis} -curve for $x > 0.33$, in good agreement with experimental observations. It can be easily proven that $\Delta E_s < \Delta E_e$, after the overlayer thickness exceeds the L_{3D} . This indicates that formation of 3D islands is energetically favourable. The dotted line in Fig. 3 is the CLT-curve derived from the single-kink Matthews model [25] applicable to a thin $\text{In}_x\text{Ga}_{1-x}\text{As}$ overlayer for comparison.

It must be emphasized that the energy consideration of strain relaxation is only a necessary but not sufficient condition. Kinetic barriers often exist and impede relaxation. For example, the Peierls–Nabarro friction force [26] is important in semiconductors at low growth temperatures. This force prevents threading dislocations or dislocation half-loops from gliding to form misfit dislocation kinks [25]. Surface diffusion, which leads to 3D islanding, also highly depends on growth temperature and growth rate [27]. We have previously measured L_{3D} as a function of the growth temperature for $0.36 \leq x \leq 1$ [19, 28, 29]. The results

showed that the CLT decreased with increased growth temperature. The temperature dependence was more significant for $0.36 \leq x \leq 0.5$. The measured L_{3D} -curve could be well fitted by the single-kink Matthews model [25] including the Peierls-Nabarro friction force, assuming that 3D islands nucleated on misfit dislocations [30].

Experimental evidences of activation of misfit dislocations in $In_xGa_{1-x}As/GaAs$ system are few. The only report [31] to our knowledge is from PL measurements on $In_xGa_{1-x}As/GaAs$ ($0.1 \leq x \leq 0.5$) single QWs annealed at 825°C for 30 minutes. The change of PL linewidth for $x \leq 0.2$ was within $\pm 1\text{ meV}$ and the peak shifted to higher energy with annealing. Hence, there were no obvious optical degradation of the annealed samples that could indicate thermal activation of misfit dislocations. In fact, both In diffusion [32] and segregation [16] at such a high annealing temperature were significant, resulting in an inclined QW potential profile and thus increased interband transition energies.

Finally, we have examined the effect of GaAs cap layer thickness on residual strain and optical quality. Typical PL spectra are shown in Fig. 1(b) for 250\AA thick $In_{0.2}Ga_{0.8}As/GaAs$ single QWs with 50 and 5000\AA thick GaAs cap layers, respectively. Photoluminescence spectra from other samples with a cap layer thickness between 50 and 5000\AA are shown elsewhere [33]. The PL spectra revealed that the transition peak shifted to a higher energy, the intensity increased and the linewidth became narrow with increasing GaAs cap layer thickness. These changes indicated improved optical quality. The TEM measurements showed a reduced dislocation density and therefore the residual strain recovered with cap layer thickness. It is interesting that misfit dislocations do exist in the sample with a 5000\AA thick cap layer having excellent optical qualities shown in Fig. 1(b). The average distance between neighbouring dislocations was found to be larger than $1\text{ }\mu\text{m}$. This implies that the optical qualities do not necessarily degrade when there exist a very small amount of dislocations. It also reveals that TEM has a better resolution than PL in measuring strain relaxation. Indium segregation was confirmed from the SIMS measurements and was found significant in the thick cap layer samples. There are several reasons for strain recovery with cap layer thickness. A thin cap layer imposes a tensile strain force on misfit dislocations. This force has a tendency to push back the existed dislocation kinks. The In segregation gives rise to reduced In contents near the interfaces. The alloy layer with a lower In content has a larger ability to accommodate the misfit. With increase of cap layer thickness, the dislocation line tension, which is counteracted by the strain force, increases [34]. The residual strain then recovers.

In conclusion, we have studied initial strain relaxation and CLT in lattice mismatched $In_xGa_{1-x}As/GaAs$ single QWs for the whole In content range. Strain relief was found by formation of misfit dislocations for $x < 0.25$ and of 3D islands for $x \geq 0.3$. Presence of misfit dislocations and 3D islands led to degradation of the optical quality. The strain relief by misfit dislocations was explained by Matthews and Blakeslee's mechanical equilibrium model. An expression for the CLT based on the onset of 3D island growth was derived using energy balance model. The L_{3D} -curve was below the L_{dis} -curve for $x > 0.33$, in good agreement with

the experimental observations. The influence of the GaAs cap layer thickness on residual strain was found to be significant. The residual strain recovered and optical qualities improved with increasing GaAs cap layer thickness.

Acknowledgements

We thank M. J. Ekenstedt, J. Y. Yao, Z. H. Lai, J. V. Thordson and U. Södervall for their valuable contributions to this work. The Swedish Board for Industrial and Technical Development (NUTEK), Swedish Research Council for Engineering Sciences (TFR) and the Swedish National Science Research Council (NFR) are acknowledged for their financial support.

References

1. Matthews, J. W., "Dislocations in Solids" (Edited by F. R. N. Nabarro) (North-Holland Publ. Co., Amsterdam 1979), p. 461.
2. Gourley, P. L., Fritz, I. J. and Dawson, L. R., *Appl. Phys. Lett.* **52**, 377 (1988).
3. Herbeaux, C., Di Persio, J. and Lefebvre, A., *Appl. Phys. Lett.* **54**, 1004 (1989).
4. Breen, K. R., Uppal, P. N. and Ahearn, J. S., *J. Vac. Sci. Technol. B7*, 758 (1989).
5. Chang, K. H., Bhattacharya, P. K. and Gibala, R., *J. Appl. Phys.* **66**, 2993 (1989).
6. Grundmann, M., Lienert, U., Christen, J., Bimberg, D., Fischer-Colbrie, A. and Miller, J. N., *J. Vac. Sci. Technol. B8*, 751 (1990).
7. Goldstein, L., Glas, F., Marzin, J. Y., Charasse, M. N. and Le Roux, G., *Appl. Phys. Lett.* **47**, 1099 (1985).
8. Yao, J. Y., Andersson, T. G. and Dunlop, G. L., *Appl. Phys. Lett.* **53**, 1420 (1988).
9. Radulescu, D. C., Schaff, W. J., Eastman, L. E., Ballingall, J. M., Ramsayer, G. O. and Hersee, S. D., *J. Vac. Sci. Technol. B7*, 111 (1989).
10. Guha, S., Madhukar, A. and Rajkumar, K. C., *Appl. Phys. Lett.* **57**, 2110 (1990).
11. Wang, S. M., Andersson, T. G., Kulakovskii, V. D. and Yao, J. Y., *Superlatt. Microstruct.* **9**, 123 (1991).
12. Yao, J. Y., Andersson, T. G. and Dunlop, G. L., *J. Appl. Phys.* **69**, 2224 (1991).
13. Ceschin, A. M. and Massies, J., *J. Cryst. Growth* **114**, 693 (1991).
14. Snyder, C. W., Orr, B. G., Kessler, D. and Sander, L. M., *Phys. Rev. Lett.* **66**, 3032 (1991).
15. Stranski, I. N. and von Krastanov, L., *Akad. Wiss. Let. Mainz. Math.-Natur. Kl. IIb*, 146 (1939).
16. Moison, J. M., Guille, C., Houzay, F., Barthe, F. and Van Rompay, M., *Phys. Rev. B40*, 6149 (1989).
17. Ekenstedt, M. J. and Andersson, T. G., *J. Vac. Sci. Technol. B9*, 1605 (1991).
18. Andersson, T. G., Chen, Z. G., Kulakovskii, V. D., Uddin, A. and Vallin, J. T., *Appl. Phys. Lett.* **51**, 752 (1987).
19. Wang, S. M., Andersson, T. G. and Ekenstedt, M. J., *Appl. Phys. Lett.* **59**, 2156 (1991).
20. Wang, S. M., Andersson, T. G., Chen, W. Q., Södervall, U. and Thordson, J., *J. Cryst. Growth* **135**, 455 (1994).
21. Kavanagh, K. L., Capano, M. A., Hobbs, L. W., Barbour, J. C., Marée, P. M. J., Schaff, W. J., Meyer, J. W., Pettit, D., Woodall, J. M., Stroscio, J. A. and Feenstra, R. M., *J. Appl. Phys.* **64**, 4843 (1988).
22. Abrahams, M. S., Blanc, J. and Buiocchi, C. J., *Appl. Phys. Lett.* **21**, 185 (1972).
23. van der Medrve, J. H., *J. Appl. Phys.* **34**, 123 (1962).
24. Matthews, J. W. and Blakeslee, A. E., *J. Cryst. Growth* **27**, 118 (1974).
25. Matthews, J. W., *J. Vac. Sci. Technol. B12*, 126 (1975).
26. Nabarro, F. R. N., "Theory of Crystal Dislocations" (Clarendon, Oxford 1967).
27. Grandjean, N. and Massies, J., *Semicond. Sci. Technol.* **8**, 2031 (1993).
28. Ekenstedt, M. J., Wang, S. M. and Andersson, T. G., *Appl. Phys. Lett.* **58**, 1991.
29. Wang, S. M., Andersson, T. G. and Ekenstedt, M. J., *Appl. Phys. Lett.* **61**, 3139 (1992).
30. Price, G. L., *Phys. Rev. Lett.* **66**, 469 (1991).
31. Elman, B., Koteles, E. S., Melman, P., Jagannath, C., Armiento, C. A. and Rothman, M., *J. Appl. Phys.* **68**, 1351 (1990).
32. Lee, J. C. and Schlesinger, T. E., *J. Vac. Sci. Technol. B5*, 1187 (1987).
33. Wang, S. M., Andersson, T. G., Lai, Z. H. and Thordson, J. V., *Semicond. Sci. Technol.* **9**, 1230 (1994).
34. Twigg, M. E., *J. Appl. Phys.* **68**, 5109 (1990).

Electron Correlation Effects in Quantum Dots

A. Matulis*

Semiconductor Physics Institute, Gostauto 11, 2600 Vilnius, Lithuania

Received May 2, 1994; accepted June 14, 1994

Abstract

Convergent renormalized perturbation series in powers of the electron-electron interaction are proposed for calculating the energy of a quantum dot. The method is illustrated by calculating the ground state of a quantum dot consisting of two electrons and the ground state of exciton in quantum dot model. Comparison between the present results and those of an exact numerical integration shows the great accuracy of the proposed method in the whole range of the electron-electron coupling constant values.

1. Introduction

Recently quantum dots containing one, two and more electrons, say quantum dot “helium” and higher “elements” respectively were created and investigated [1–3]. The exchange and correlation effects were shown to be of great importance [4]. The theoretical investigation of the correlation effects is mainly based on the straightforward numerical Hamiltonian diagonalization in the many particle function space [4, 5]. It is a rather laborious procedure.

We propose to treat the electron system in the quantum dot in more simple way making use of the expansion in a power series in electron-electron (ee) interaction. The correlation effects are known to manifest themselves in the second order of that expansion already. Actually, the expansion is in powers of the dimensionless ee-interaction coupling constant $\lambda = a_0/a_B$ where a_0 is the characteristic quantum dot dimension, $a_B = \epsilon\hbar^2/m^*e^2$ is the Bohr radius in which ϵ is the static dielectric constant of the medium in which the electrons are located in and m^* is the electron effective mass. Thus, the effective coupling constant can be easily varied by changing the confining potential. Typically, values $\lambda \geq 2$ are of interest experimentally.

The primitive λ -expansion is only useful in the case of small coupling constant, i.e. for very narrow quantum dots. In order to cover the whole range of λ values we constructed the renormalized λ -expansion as was previously used in the case of the nonparabolic oscillator problem (see, for example, [6]). The renormalized λ -expansion is constructed by means of two orders of simple λ -expansion, the asymptotic expansion for large λ values and taking into account the scaling properties of the Schrödinger equation.

For illustrative purposes we considered the ground state of the two electron system and electron-hole system in a parabolic quantum dot and compared the obtained results with the exact numerical solution. An agreement within 1% was found for all λ values.

2. Formulation of the problem

Consider N electrons in a two-dimensional quantum dot described by the following Schrödinger equation

$$\left\{ \sum_{i=1}^N \left[-\frac{\hbar^2}{2m^*} \nabla_i^2 + \frac{1}{2} m^* \omega_0^2 r_i^2 \right] + \frac{e^2}{\epsilon} \sum_{i < j} \frac{1}{|r_i - r_j|} - E \right\} \Psi = 0 \quad (2.1)$$

where ω_0 is the confinement frequency parameter. For future purposes we will introduce dimensionless units by making the following transformation: 1) $r \rightarrow a_0 r$, where $a_0 = \sqrt{\hbar/m^*\omega_0}$, 2) the energy will be measured in $\hbar\omega_0$ units. Using these dimensionless notations we obtain the following eigenvalue problem

$$\left[-\frac{1}{2} \sum_{i=1}^N (\nabla_i^2 - r_i^2) + \lambda \sum_{i < j} \frac{1}{|r_i - r_j|} - E(\lambda) \right] \Psi = 0. \quad (2.2)$$

Making use of the perturbation methods the eigenvalue can be expanded into power series. Let us suppose that we know that expansion

$$E(\lambda) = E_0 + \lambda E_1 + \lambda^2 E_2. \quad (2.3)$$

The coefficients E_0 , E_1 and E_2 can be expressed by unperturbed one electron energies in the parabolic quantum dot

$$e(i) = 1 + |m_i| + 2n_i \quad (2.4)$$

and the Coulomb matrix elements

$$\begin{aligned} V_{i, j; j', i'} &= \delta_{s_i, s_{i'}} \delta_{s_j, s_{j'}} \int d^2 r_1 \int d^2 r_2 \frac{1}{|r_1 - r_2|} \\ &\times \phi^*(i|r_1) \phi^*(j|r_2) \phi(j'|r_2) \phi(i'|r_1) \\ &= \delta_{s_i, s_{i'}} \delta_{s_j, s_{j'}} \delta_{l+m_{i'}, m_i} \delta_{m_{j'}, l+m_j} \\ &\times R_l(m_i n_i, m_{i'} n_{i'}) |m_j n_j, m_{j'}, n_{j'}), \end{aligned} \quad (2.5)$$

$$R_l(m_1 n_1, m_2 n_2 | m_3 n_3, m_4, n_4)$$

$$\begin{aligned} &= \int_0^\infty r dr \int_0^\infty r' dr' \int_0^{2\pi} d\varphi \frac{e^{il\varphi}}{\sqrt{r^2 + r'^2 - 2rr' \cos(\varphi)}} \\ &\times R(m_1, n_1 | r) R(m_2, n_2 | r) R(m_3, n_3 | r') R(m_4, n_4 | r') \end{aligned} \quad (2.6)$$

Here symbol $i = \{m_i, n_i, s_i\}$ stands for the set of one electron quantum numbers.

* Electronic mail: matulis@kktlpf.lt

The matrix element should be calculated with the oscillator functions

$$\phi(i| \mathbf{r}) = \frac{1}{\sqrt{2\pi}} e^{im_i\varphi} R(m_i, n_i | r) \quad (2.7)$$

which can be expressed via associated Laguerre polynomials

$$R(m, n | r) = \sqrt{\frac{2n!}{(m+n)!}} r^{|m|} e^{-r^2/2} L_n^m(r^2). \quad (2.8)$$

3. Renormalized series

It is obvious that the two orders of λ -expansion (2.3) could be useful for small λ values only. Now we shall try to improve the convergence of the obtained λ -expansion by some renormalization procedure. We construct the renormalized λ -expansion following the procedure described in [6]. The main idea of this renormalization is a scaling transformation of the Schrödinger equation which enables to transform the large λ value problem into another problem which can be solved.

Instead of considering the basic eigenvalue problem (2.2) we will study the generalized eigenvalue problem as is described by the following Schrödinger equation

$$\left[\sum_{i=1}^N \left(-\frac{\xi^2}{2} \nabla_i^2 + \frac{1}{2} r_i^2 \right) + \sum_{i < j} \frac{\lambda}{|\mathbf{r}_i - \mathbf{r}_j|} - E(\xi, \lambda) \right] \Psi = 0. \quad (3.1)$$

The difference between this equation and the basic equation is the additional parameter ξ which is included in the kinetic part of the Hamiltonian. As a consequence the eigenvalue of this new problem will be a function of two parameters: λ and ξ . It is clear that one can obtain the eigenvalue of the basic problem as

$$E(\lambda) = E(1, \lambda). \quad (3.2)$$

After performing the coordinate transformation $\mathbf{r} \rightarrow \sqrt{\xi} \mathbf{r}$ one notices that the eigenvalue satisfies the following scaling condition

$$E(\xi, \lambda) = \xi E(1, \lambda \xi^{-3/2}). \quad (3.3)$$

This condition connects the eigenvalues on the trajectory in the parameter $\xi\lambda$ -plane $\lambda = \lambda_0 \xi^{3/2}$. Although according to eq. (3.2) we need the eigenvalue on the line $\xi = 1$ we can use the other suitable value on the above trajectory. As a consequence we notice that the problem of large λ can be reduced to the problem of small ξ . The parameter ξ^2 can be considered as a reciprocal electron mass. Therefore, the limiting case of $\xi \rightarrow 0$ is the quasi-classical approximation of the quantum dot problem which has been studied in [8, 9].

The quasi-classical approximation can be obtained by noticing that in the case with very large electron mass, i.e. when $\xi \rightarrow 0$, the electrons should be located near the minimum of the potential. So, calculating the derivative and equating it to zero

$$\frac{\partial}{\partial \mathbf{r}_k} \left(\sum_{i=1}^N \frac{1}{2} r_i^2 + \sum_{i < j} \frac{\lambda}{|\mathbf{r}_i - \mathbf{r}_j|} \right) = 0 \quad (3.4)$$

one can show that the minimum is located at $r_k \sim \lambda^{1/3}$ and consequently $E(\lambda) \sim \lambda^{2/3}$. Therefore we shall assume that the

following asymptotic eigenvalue expansion holds

$$E(\lambda) = c_0 \lambda^{2/3} + c_1. \quad (3.5)$$

Now we shall expand the eigenvalue $E(\xi, \lambda)$ into a ξ and λ power series choosing the lowest possible polynomial

$$\begin{aligned} E(1 - \beta, \beta) &= b_0 + b_1 \beta + b_2 \beta^2 + b_3 \beta^3 + b_4 \beta^4 \\ &= (b_0 + b_1 + b_2 + b_3 + b_4) \\ &\quad - (b_1 + 2b_2 + 3b_3 + 4b_4)(1 - \beta) + \dots, \end{aligned} \quad (3.6)$$

The coefficients of that expansion should be determined by adjusting that expansion to the eigenvalue λ -series (2.3) and to the quasi-classical asymptotic solution (3.5). The only problem we are faced with is the mapping of the λ parameter in the $\xi\lambda$ -plane. Following Ref. [6] we introduce the mapping onto the straight line

$$\begin{aligned} \xi &= 1 - \beta, \\ \lambda &= \beta. \end{aligned} \quad (3.7)$$

It is supposed that $0 < \beta < 1$.

So, in the case $\lambda \rightarrow 0$ we use the expansion

$$\begin{aligned} E(1 - \beta, \beta) &= (1 - \beta) E\left(1, \frac{\beta}{(1 - \beta)^{3/2}}\right) \\ &= (1 - \beta) \left\{ E_0 + E_1 \frac{\beta}{(1 - \beta)^{3/2}} + E_2 \frac{\beta^2}{(1 - \beta)^3} \right\} \\ &= E_0 + (E_1 - E_0)\beta + \left(\frac{E_1}{2} + E_2 \right) \beta^2 + \dots \end{aligned} \quad (3.8)$$

and in the case $\beta \rightarrow 1$, the expansion

$$\begin{aligned} E(1 - \beta, \beta) &= (1 - \beta) E\left(1, \frac{\beta}{(1 - \beta)^{3/2}}\right) \\ &= (1 - \beta) \left(c_0 \frac{\beta^{2/3}}{1 - \beta} + c_1 \right) \\ &= c_0 + (c_1 - \frac{2}{3} c_0)(1 - \beta) + \dots. \end{aligned} \quad (3.9)$$

Now comparing eqs (3.8), (3.9) and (3.6) we obtain the expansion coefficients

$$\begin{aligned} b_0 &= E_0, \\ b_1 &= E_1 - E_0, \\ b_2 &= E_1/2 + E_2, \\ b_3 &= -E_0 - 4E_1 - 2E_2 + 10c_0/3 + c_1, \\ b_4 &= E_0 + 5E_1/2 + E_2 - 7c_0/3 - c_1 \end{aligned} \quad (3.10)$$

and finally arrive at the renormalized series for the considered eigenvalue $E^{\text{ren}}(\lambda)$

$$E^{\text{ren}} = (1 - \beta)^{-1} (b_0 + b_1 \beta + b_2 \beta^2 + b_3 \beta^3 + b_4 \beta^4), \quad (3.11)$$

$$\lambda = \frac{\beta}{(1 - \beta)^{3/2}}. \quad (3.12)$$

presented in a parametric way.

4. Exciton in a quantum dot

The developed renormalization technique can be also applied to the exciton in a quantum dot problem. Actually, according to [7] the exciton can be described by the following dimensionless Schrödinger equation:

$$\left[-\frac{1}{2}(\nabla_e^2 + k_0 \nabla_h^2 - r_e^2 - k_1 r_h^2) - \lambda \frac{1}{|r_e - r_h|} - E(\lambda) \right] \Psi = 0 \quad (4.1)$$

where k_0 stands for electron and hole effective mass ratio and k_1 stands for the ratio of squared hole and electron confinement potential frequencies.

The eigenvalue λ -expansion of Eq. (4.1) can be expressed via the same Coulomb matrix elements with differently scaled electron and hole wave functions. In fact it differs from expansion (2.3) by λ sign only. The asymptotic behaviour of exciton equation, however, is quite different. In the case when $\lambda \rightarrow \infty$ electron-hole system turns into the free exciton which is known to have the following asymptotic eigenvalue λ -dependence:

$$E(\lambda) = d_0 \lambda^2 + d_1. \quad (4.2)$$

Due to the above asymptotic behaviour, for constructing the renormalized λ -expansion the following generalized eigenvalue problem should be used:

$$\left\{ -\frac{1}{2}[\nabla_e^2 + k_0 \nabla_h^2 - \xi^2(r_e^2 + k_1 r_h^2)] - \lambda \frac{1}{|r_e - r_h|} - E(\xi, \lambda) \right\} \Psi = 0 \quad (4.3)$$

Performing the coordinate transformation $r \rightarrow r/\sqrt{\xi}$ we get the eigenvalue scaling equation

$$E(\xi, \lambda) = \xi E(1, \lambda \xi^{-1/2}). \quad (4.4)$$

Now, following the procedure described in the previous section, i.e. expanding eigenvalue according to eq. (3.6) and comparing it with eqs (2.3) and (4.2) with the aid of eq. (4.4) we get the final renormalized expansion:

$$E^{\text{ren}} = (1 - \beta)^{-1}(b_0 + b_1 \beta + b_2 \beta^2 + b_3 \beta^3 + b_4 \beta^4), \quad (4.5)$$

$$\beta = \frac{\lambda^2}{2} (\sqrt{1 + 4/\lambda^2} - 1) \quad (4.6)$$

where the coefficients are defined as follows

$$\begin{aligned} b_0 &= E_0, \\ b_1 &= E_1 - E_0, \\ b_2 &= E_2 - E_1/2, \\ b_3 &= 2d_0 + d_1 - E_0 - 2E_1 - 2E_2, \\ b_4 &= E_0 + 3E_1/2 + E_2 - d_0 - d_1, \end{aligned} \quad (4.7)$$

5. Numerical illustration and conclusions

In order to illustrate the accuracy of the above λ -expansion we applied that technique to the system of two electrons in

a parabolic quantum dot for which there exist exact results to which we may compare our approximate results (see, for example [10]). Indeed introducing center-of-mass coordinates $R = (r_1 + r_2)/2$ and relative coordinates $r = r_1 - r_2$ one can separate the variables and reduce the two electron problem (2.2) into two one-particle equations which can be easily integrated.

Now let us consider the corresponding λ -expansion. Knowing the ground state configuration $m_1 = m_2 = n_1 = n_2 = 0$ and taking eq. (2.4) into account we have $E_0 = 2$. The first order correction can be expressed as

$$E_1 = R_0(00, 00 | 00, 00) = \sqrt{\frac{\pi}{2}}. \quad (5.1)$$

The second order correction is

$$E_2 = - \sum'_{l, n_1, n_2} \frac{R_l^2(l n_1, 00 | 00, l n_2)}{2(n_1 + n_2 + |l|)} = -0.345. \quad (5.2)$$

The prime indicates that the term with $l = n_1 = n_2 = 0$ should be omitted.

We should notice that equating n_1 or n_2 to zero in eq. (5.2) one gets the second order Hartree-Fock correction. For the ground state it is $E_2^{\text{HF}} = -0.105$. Comparing it with E_2 we see that the Hartree-Fock contribution covers only the half of the second order correction what confirms that the correlation energy is important.

The asymptotic expansion coefficients (3.5) were calculated by finding the minimum of the potential and approximating it by the corresponding parabolic quantum well. Those coefficients are $c_0 = 3/2^{4/3}$ and $c_1 = \sqrt{3}/2$.

Applying the considered technique to the model exciton problem of electron and hole having the same masses and confining frequencies we used the same λ -expansion (with the opposite λ sign). We also used the asymptotic expansion coefficients $d_0 = -1$ and $d_1 = 1$. The first coefficient follows from free exciton consideration and the second one is caused by the center-of-mass motion.

Making use eqs (3.10) and (4.7) the renormalized expansion coefficients were calculated. All results are shown in Fig. 1 for two electron system case and in Fig. 2 for exciton case. Notice that the simple λ -expansion can be successfully

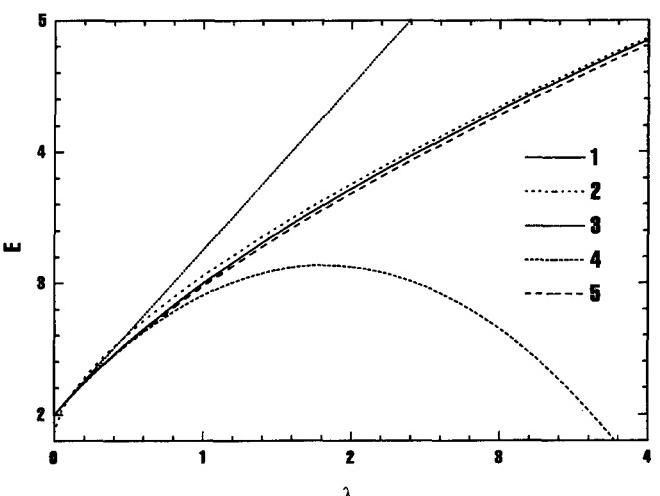


Fig. 1. The two electrons ground state energy versus ee-interaction plot: 1 – exact numerical result, 2 – asymptotic expansion, 3 – first order λ -expansion, 4 – second order λ -expansion, 5 – renormalized λ -expansion.

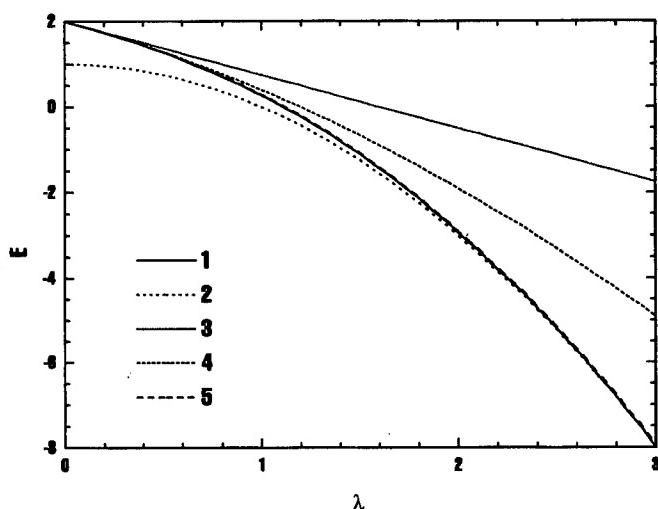


Fig. 2. The exciton ground energy versus eh-interaction plot. Legend is the same as in Fig. 1.

used in the $\lambda < 1$ region, but for $\lambda \geq 1$ it diverges away from the exact result. It is seen that in both cases the agreement of the renormalized expansion with the exact result is better than 1% for all λ -values. The same coincidence of the renor-

malized expansion with the exact result were obtained for excited states too.

Acknowledgements

The author would like to express his sincere gratitude to the Baltic Scholarship Scheme of the Nordic Council of Ministers.

References

1. Sikorski, Ch. and Merkt, U., Phys. Rev. Lett. **62**, 2164 (1989).
2. Lorke, A., Kotthaus, J. P. and Ploog, K., Phys. Rev. Lett. **64**, 2559 (1990).
3. Ashoori, R. C. et al., Phys. Rev. Lett. **68**, 3088 (1992).
4. Pfankuche, D., Gudmundsson, V. and Maksym, P., Phys. Rev. **B47**, 2244 (1993).
5. Maksym, P. A. and Chakraborty, T., Phys. Rev. Lett. **65**, 108 (1990).
6. Fernández, F. M., Artega, G. A., Maluendes, S. A. and Castro, E., Physics Letters **103A**, 19 (1984).
7. Bockelmann, U., Phys. Rev. **B48**, 17637 (1994).
8. Bolton, F. and Rössler, U., Superlatt. Microstruct. **13**, 139 (1993).
9. Bedanov, V. M. and Peeters, F. M., Phys. Rev. **B49** (1994).
10. De Groot, Jean-Jacques S., Hornos, J. E. M. and Chaplik, A. V., Phys. Rev. **B49**, 12773 (1992).

Enhancement of the g -Factor and Spin-Density Wave State in a Confined 2DEG in the Quantum Hall Regime

Vidar Guðmundsson* and Gunnar Pálsson

Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland

Received May 19, 1994; accepted June 14, 1994

Abstract

We investigate the spin splitting of the Landau bands (LB's) in a confined two-dimensional electron gas (2DEG) using the Hartree-Fock approximation (HFA) for the mutual Coulomb interaction of the electrons. The exchange term of the interaction causes a large splitting of the spin levels of a LB whenever the chemical potential lies between them. These oscillations of the splitting with the filling factor of the LB's are conveniently interpreted as an oscillating enhancement of the effective g -factor, g^* . The reduction of g^* when a LB is becoming completely filled is accompanied by a spontaneous formation of a static spin-density wave state whose details depend on the system size and temperature.

1. Introduction

The enhancement of the effective g -factor, g^* , of a 2DEG in the quantum Hall regime has been studied experimentally and theoretically by several researchers [1]. Ando and Uemura [2] presented a model of an unbounded 2DEG where the broadening of the Landau levels due to impurity scattering is treated in the self-consistent Born approximation (SCBA). The dielectric function is calculated with the inclusion of the lowest order exchange energy of the screened Coulomb interaction in the self-energy of the electrons. For a strong magnetic field the overlapping of Landau levels with different indices is neglected. In addition to the enhancement of g^* the exchange interaction can lead to the spontaneous formation of spin-density [3] or charge-density waves [4, 5] in two-dimensional electron systems.

The instability of the spin-density wave in wide parabolic quantum wells has been studied by Brey and Halperin using a modified HFA with a point-contact exchange interaction. They find a divergence of the electric susceptibility in the presence of a magnetic field of intermediate strength parallel to the quantum well and an infinitesimal fictitious magnetic field perpendicular to the quantum well [6]. This spin-density wave state has a wavevector along the quantum well parallel to the intermediate magnetic field and occurs only when the quantum well is wide enough and the exchange interaction has a strength larger than a critical value. Kempa *et al.* have investigated the spontaneous polarization of an array of quantum dots into a ferroelectric or anti-ferroelectric state [7].

Here we shall study the spin splitting of LB's, the oscillations of g^* , and the formation of spin-density wave state in

a strictly two dimensional mesoscopic 2DEG system in a perpendicular magnetic field of intermediate strength. We thus observe a spontaneous spatial dependent polarization of the 2DEG within one mesoscopic system of the size of a quantum dot. We investigate the effects of the finite size of the system on the LB's in the absence of any impurity scattering of the electrons. The system size is chosen to be of the order of several magnetic lengths $l = \sqrt{\hbar c/(eB)}$. The LB's in the center of the system do then approach flat Landau levels indicating that an electron in the center does not feel the boundary. We are thus able to observe effects caused by the finite size of the system and the bulk character of some of the electron states together with the fine interplay of these two effects characteristic to mesoscopic systems. To accomplish these goals the mutual Coulomb interactions of the exactly two-dimensional electrons in several LB's are included in the Hartree-Fock approximation at finite temperature. The effects of the exchange interaction on the screening of a simple Coulomb impurity by the 2DEG have been discussed elsewhere [8].

2. Model

We consider N_s strictly two-dimensional electrons to model qualitatively a real heterostructure where the 2DEG is confined to the lowest electrical subband. The 2DEG is confined to a disk of radius R in the 2D-plane by a potential step

$$V_{\text{conf}}(r) = U_0 \left[\exp \left(\frac{R-r}{4\Delta r} \right) + 1 \right]^{-1}, \quad (1)$$

where $\Delta r = 22 \text{ \AA}$. To ensure charge neutrality of the system a positive back-ground charge n_b resides on the disk

$$n_b(r) = \bar{n}_s \left[\exp \left(\frac{r-R}{\Delta r} \right) + 1 \right]^{-1}, \quad (2)$$

with the average electron density of the system given by $\bar{n}_s = N_s/(\pi R^2)$. In the Hartree-Fock approximation the state of each electron is described by a single-electron Schrödinger equation

$$\{H^0 + V_H(r) + V_{\text{conf}}(r)\}\psi_\alpha(r) - \int d^2r' \Delta(r, r') \psi_\alpha(r') = \varepsilon_\alpha \psi_\alpha(r) \quad (3)$$

* vidar@raunvis.hi.is

for an electron moving in a Hartree potential

$$V_h(r) = \frac{e^2}{\kappa} \int d^2r' \frac{n_s(r') - n_b(r')}{|r - r'|} \quad (4)$$

and a nonlocal Fock potential with

$$\Delta(r, r') = \frac{e^2}{\kappa} \sum_{\beta} f(\epsilon_{\beta} - \mu) \frac{\psi_{\beta}^*(r')\psi_{\beta}(r)}{|r - r'|}, \quad (5)$$

where $f(\epsilon_{\beta} - \mu)$ is the Fermi distribution at the finite temperature T . $n_s(r)$ is the electron density

$$n_s(r) = \sum_{\alpha} |\psi_{\alpha}(r)|^2 f(\epsilon_{\alpha} - \mu), \quad (6)$$

with the chemical potential μ . The label α represents the radial quantum number n_r , the angular quantum number M , and the spin quantum number $s = \pm \frac{1}{2}$. H^0 is the single particle Hamiltonian for one electron with spin in a constant perpendicular external magnetic field [9, 10]. A Landau band index n can be constructed from the quantum numbers n_r and M as $n = (|M| - M)/2 + n_r$. The Landau levels of H^0 with energy $E_{n, M, s} = \hbar\omega_c(n + \frac{1}{2}) + sg^*(\mu_B/\hbar)B$ are degenerate with respect to M with the degeneracy $n_0 = (2\pi l^2)^{-1}$ per spin orientation. μ_B is the Bohr magneton ($e\hbar/2mc$). The filling factor of the Landau levels is defined by $v = \bar{n}_s/n_0 = 2\pi l^2 \bar{n}_s$. It assumes even integer values when an even number of spin Landau levels is filled. The cyclotron frequency is given by $\omega_c = eB/(mc)$. Hartree-Fock energy spectrum ϵ_{α} and the corresponding wave functions are now found by solving (3)–(6) iteratively in the basis [10–12] of H^0 . The chemical potential μ is recalculated in each iteration in order to preserve the total number of electrons N_s . The number of basis functions used in the diagonalization is chosen such that a further increase of the subset results in an unchanged density $n_s(r)$.

3. Results

The calculations are carried out with GaAs parameters: $m^* = 0.067m_e$, $\kappa = 12.4$, and $g^* = -0.44$. The occupation of the LB's is varied by changing N_s at a constant strength of the magnetic field $B = 3.0$ T. Since the radius of the system $R \geq 1000$ Å is much larger than the magnetic length $l \approx 148$ Å and the effective Bohr radius $a_0^* \approx 97.9$ Å we can use the average filling factor $v = 2\pi l^2 \bar{n}_s$ to describe the occupation of the lower bands in the interior of the system. The cyclotron energy $\hbar\omega_c \approx 5.2$ meV, so a sufficient height of the confining potential is $U_0 = 80$ meV in order to include several LB's in the calculation. For $B = 3.0$ T the bare spin splitting of the LB's ($g^*\mu_B/\hbar)B \approx 0.076$ meV is much smaller than their separation $\hbar\omega_c$ and corresponds to the thermal energy $k_B T$ at $T \approx 0.9$ K.

Figure 1 shows the energy spectra of the electrons for 4 different values of N_s at $T = 10$ K. For $N_s = 48$ the chemical potential μ is located between the two lowest almost spin degenerate LB's. When N_s increases states of the LB with quantum numbers ($n = 1, s = +1/2$) become occupied and drop below μ until around $N_s = 62$ all the bulk states of that LB are occupied and are well separated from the empty LB with the same Landau level index n but opposite spin $s = -1/2$. Further increase in N_s brings the bulk states of

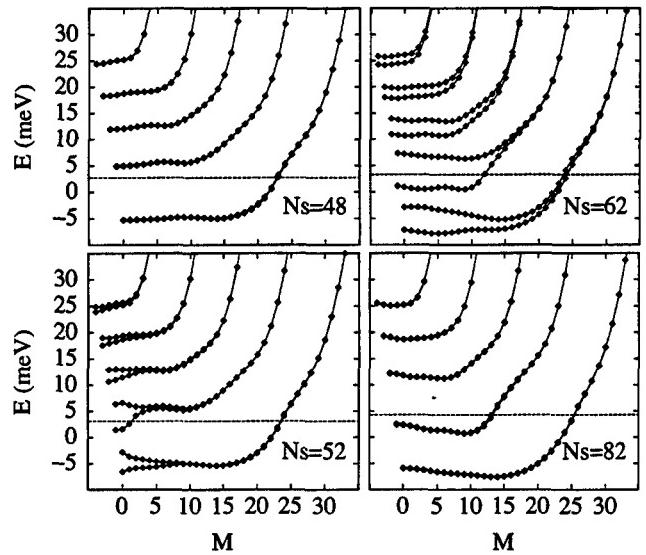


Fig. 1. Energy spectra and chemical potential μ (horizontal dashed line). The total number of electrons N_s is indicated in each subfigure. $T = 10$ K, $R = 1000$ Å, $U_0 = 80$ meV, and $B = 3.0$ T. GaAs bulk parameters: $m^* = 0.067m_e$, $\kappa = 12.4$, $g^* = -0.44$.

this second LB all below μ drastically reducing the spin splitting. The root mean square g-factor, g_{rms}^* for the lowest three LB's is seen in Fig. 2, where the splitting of the LB's has been interpreted as the normal Zeeman spin splitting ($g^*\mu_B/\hbar)B$ with an effective g-factor. g_{rms}^* clearly takes on the bare value of 0.44 for even integer filling factor v , when an equal number of spin up and down states are occupied. On the other hand g_{rms}^* peaks whenever v assumes values in the neighborhood of odd integers, then μ lies between two spin LB's with the same n . The unequal number of spin up and down states then leads to different exchange energies for the two different spin directions [2]. The enhancement of the g-factor is well known from Shubnikov-de Haas experiments and the concurrent splitting of all the Landau levels has been established in optical measurements [13].

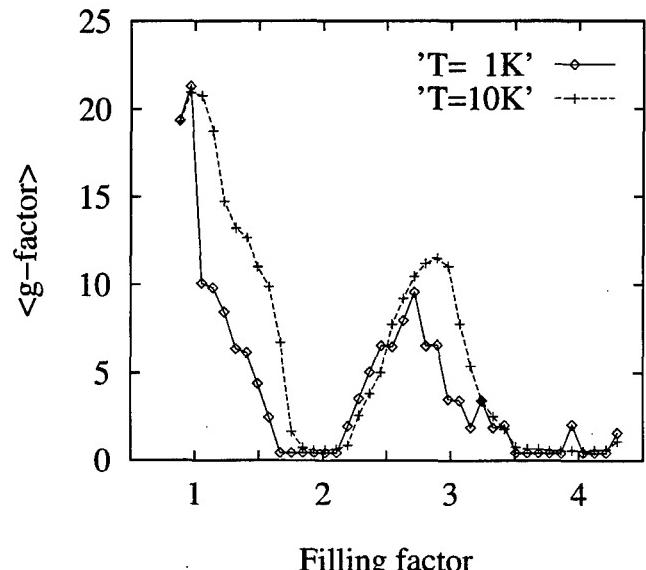


Fig. 2. g_{rms}^* as a function of the average filling factor v for $T = 1$ K (solid), and 10 K (dashed). Other parameters are as in Fig. 1.

We have discovered quite an interesting phenomenon whenever N_s is increased such that v is changing from an odd integer to an even integer value, that is when the spin splitting of the LB's is closing down. This situation is displayed in Fig. 3 for the lowest LB. This apparent twisting of the LB's is repeated for a narrower range of v or N_s for μ in higher LB's. Figure 4 showing the total density $n_s(r)$ and the electron density for each spin direction separately confirms that a static spin-density wave state forms spontaneously in the system concurrently with the twisting of the LB's.

An important difference of the present spin-density wave state in the 2D plane to the one parallel to B investigated by Brey and Halperin [6] is the fact that the wavelength of the present modulation varies strongly with v . This is caused by the strong dependence of the effective interaction, or the screening, in the 2D plane on v [10, 14–16]. One might expect the enhancement of g^* and the spin-density wave state to interfere with the formation of incompressible regions of the 2DEG seen in the model of McEuen et al. [16].

To exclude numerical deficiencies we have tested the stability of the spin-density structures by increasing the number of basis states included in the numerical calculation. The exact shape and formation of the spin-density waves state does depend on the size of the system emphasizing that we are observing a confined spin-density wave state here [8]. In larger systems the spin-density wave occurs spontaneously together with a charge-density wave that dominates for stronger magnetic fields. Brey and Halperin can inhibit the formation of a charge-density wave state in their model by keeping the exchange interaction not too strong [6]. Calculations at a much higher temperature ($T = 100$ K) show no sign of any spin splitting since the thermal energy is then much larger than the Zeeman energy and of the same order as the separation of the LB's, $\hbar\omega_c$. The spin-density wave is better developed and exists for a larger region of v for 10 K than 1 K. At the higher temperature the occupation of the different spin orientations is almost the same, such that small perturbations to the Landau levels can have large

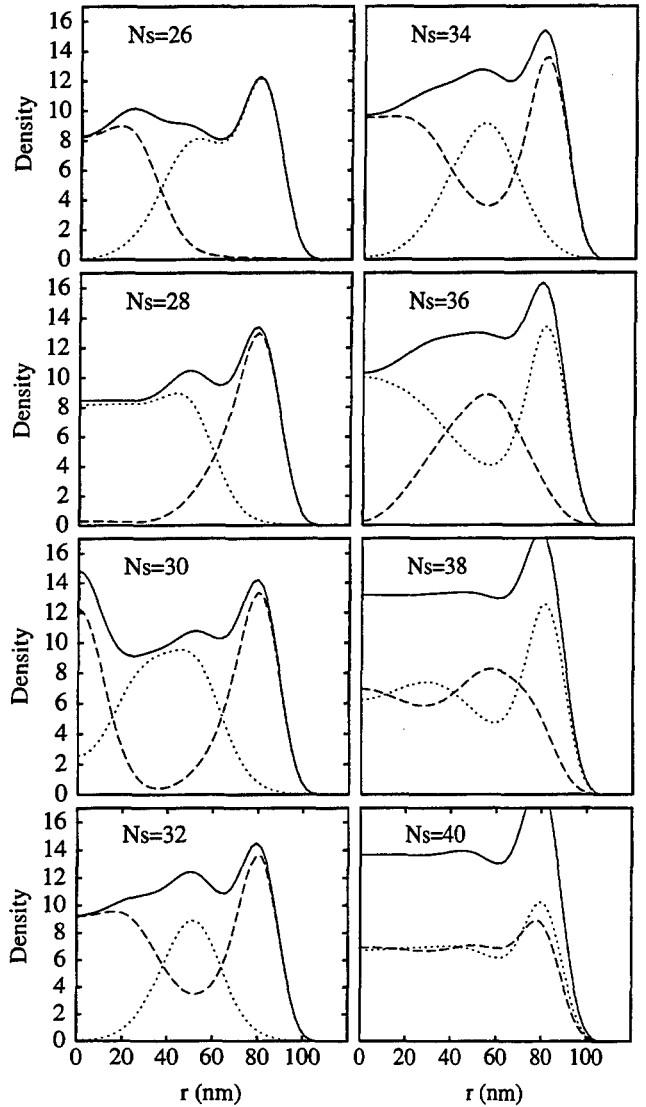


Fig. 4. The electron density $n_s(r) \times 10^4 \text{ nm}^{-2}$ for both spin directions (solid), spin up $s = +1/2$ (dashed), and spin down $s = -1/2$ (dotted) as functions of the radius r of the system. Other parameters are as in Fig. 1.

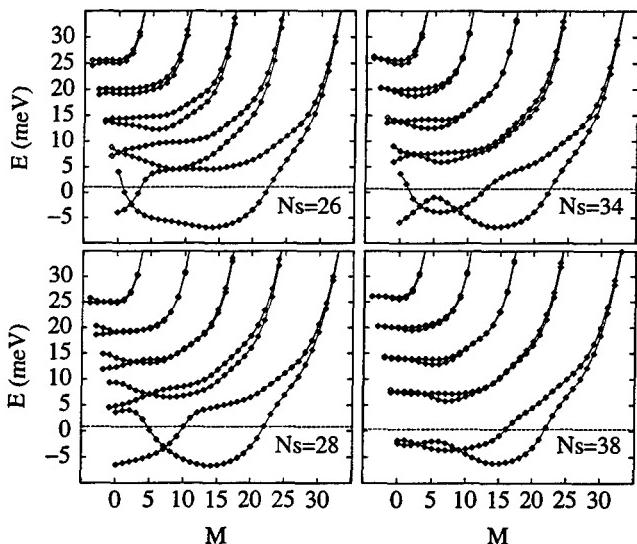


Fig. 3. Energy spectra and chemical potential μ (horizontal dashed line). The total number of electrons N_s is indicated in each subfigure. Other parameters are as in Fig. 1.

effects in the self-consistent evaluation of the LB's. The exchange interaction reduces the Coulomb repulsion between electrons of opposite spin orientation, and thus favors the spontaneous formation of a spin-density wave. The reduced dimensionality of the electron gas strengthens the electron-electron interaction and gives the spatial ordering of the spins an extra weight in the competition with the effects of the magnetic field that tend to order the spins all in the same direction. At much higher temperatures thermal excitations make states in neighboring LB's accessible for rearrangement of the 2DEG to neutralize any forces that might otherwise lead to splitting of LB's or spatial variations of the spin or charge densities.

4. Summary

In a mesoscopic system of a confined 2DEG we have been able to demonstrate both bulk effects and phenomena caused by the finite size of the system, in the absence of any impurity scattering of the electrons. The 2D system is large

enough so that the LB's approach flat Landau levels for low values of the angular quantum number M . This can be interpreted as the formation of 2D bulk states inside the system. The ensuing singular density of states together with the exchange interaction causes the well known oscillations of the energy separation of the LB's with the same Landau level index n but opposite spin orientations as a function of the filling factor v . Here we have seen that the enhancement of g^* occurs not only in the LB where μ is located but in all the LB's included in the model. Similar behavior has been established in optical measurements of a 2DEG by Kukushkin [13]. In a large system of 2D electrons the state of homogeneous density becomes unstable and a charge-density wave forms spontaneously [4, 5, 17]. In a small system the formation of such a wave is energetically unfavorable, but becomes possible with an increasing system size. Here we observe the spontaneous formation of concentric circular regions of different spin phases whenever the spin splitting of the LB's is closing down with an increasing v at a low temperature. The shape of this spin-density wave depends on the size of the system, the temperature, and the filling factor v , such that the wavelength decreases as v approaches an even integer. With an increasing size of the system the spin-density wave acquires properties of a charge-density wave. In the extreme quantum limit (when $v \rightarrow 0$) no such phase separation is observed in the Hartree-Fock approximation and only one spin state is occupied.

Whether this formation of a spin-density wave state in the HFA for finite systems is physical has to be investigated in higher order approximations. As the Hartree-Fock approximation does not describe the broadening of nearly degenerate Landau levels due to the Coulomb interaction correctly, it can be expected that this approximation overemphasizes the importance of the exchange interaction between the electrons, so one might speculate whether approximations of higher order would decrease the enhancement of the spin splitting or alter the occurrence of the spin-density wave state.

In order to limit the approximations needed for the electron-electron interaction we have chosen the system to be strictly two-dimensional, this certainly precludes quanti-

tative comparison with experimental results and the qualitative validity of the model in the limit of a 2DEG with a finite thickness has to be established. Models of 2DEG in quantum dots of finite thickness seem to indicate the qualitative applicability of such calculations as far as the electric subbands are well separated [18].

The present results give an indication what to expect in optical measurements of the properties of a 2DEG in lightly doped quantum wells and heterostructures, but calculations of the cyclotron resonance or the plasmonic structure of the model are necessary in order to compare with experimental data.

Acknowledgements

This research was supported in part by the Icelandic Natural Science Foundation.

References

1. Ando, T., Fowler, A. B. and Stern, F., Rev. Mod. Phys. **54**, 437 (1982).
2. Ando, T. and Uemura, Y., J. Phys. Soc. Japan **37**, 1044 (1974).
3. Grüner, G., Rev. Mod. Phys. **66**, 1 (1994).
4. Gerhardts, R., Phys. Rev. **B24**, 1339 (1981).
5. Yoshioka D. and Lee P., Phys. Rev. **B27**, 4986 (1983).
6. Brey, L. and Halperin, B., Phys. Rev. **B40**, 11634 (1989).
7. Kempa, K., Broido, D. and Bakshi, P., Phys. Rev. **B43**, 9343 (1991).
8. Gundmundsson, V. and Pálsson, G., Phys. Rev. **B49**, 13712 (1994).
9. Landau, L. and Lifshitz, E., "Quantum Mechanics" (Pergamon Press, London 1958).
10. Gundmundsson, V., Solid State Comm. **74**, 63 (1990).
11. Gundmundsson, V., in: "Proceedings of the NATO Advanced Research workshop on Semiconductor Microstructures, May 1989, Venezia" (Edited by A. F. G. Fasol and P. Lugi) (Plenum Press, New York 1989), p. 517.
12. Pfankuche, D., Gundmundsson, V. and Maksym, P., Phys. Rev. **B47**, 2244 (1983).
13. Kukushkin, I., Meshkov, S. and Timofeev, Sov. Phys. Usp. **31**, 511 (1988).
14. Wulf, U., Gundmundsson, V. and Gerhardts, R., Phys. Rev. **B38**, 4218 (1988).
15. Labbé, J., Phys. Rev. **B35**, 1373 (1988).
16. McEuen, P. et al., Phys. Rev. **B45**, 11419 (1992).
17. MacDonald, A., Phys. Rev. **B30**, 4392 (1984).
18. Kumar, A., Laux, S. and Stern, F., Phys. Rev. **B42**, 5166 (1990).

Dissipation in the Quantum Hall Effect by Transverse Circulation of Electrons

E. B. Hansen

Oersted Laboratory, Niels Bohr Institute, Universitetsparken 5, DK-2100 Copenhagen, Denmark

Received May 3, 1994; accepted June 14, 1994

Abstract

In a mid-plateau situation there exists a reversible uphill flow (in the direction of the Hall field E_y) of the electron liquid with the current density $J_{\text{liq.}}$. Inelastic processes cause excitations to be scattered downhill with the current density $J_{\text{exc.}}$, which satisfies the condition $J_{\text{exc.}} + J_{\text{liq.}} = 0$. The conversion at the edge of the reversible uphill flow to the dissipative downhill flow is mediated by the edge states. It is shown that transverse circulation of electrons implies the temperature independence of the ratio $\Delta\rho_{xy}/\rho_{xx}^{\min}$, where ρ_{xx}^{\min} is the minimal value of the longitudinal resistivity at the given temperature T , and $\Delta\rho_{xy} = \rho_{xy}(T) - \rho_{xy}(0)$.

1. Introduction

Superimposed on the longitudinal transport current I there is a transverse circulation of electrons. This circulation consists of two opposite flows: a reversible uphill flow of the electron liquid and an irreversible downhill flow of excitations. In the uphill flow electrons move in the direction of the Hall field E_y , so this flow implies the storage of potential energy. The driving field for the uphill flow is the longitudinal electric field E_x . The downhill flow of excitations can take place only via phonon emission. Therefore, in a steady state, this flow dissipates potential energy at the same rate as it is stored by the uphill flow. In ordinary conduction the work done by the driving field is initially stored as kinetic energy, which, subsequently, is dissipated at the same rate as it is stored. It is one of the remarkable features of the quantum Hall effect that it is potential energy rather than kinetic energy which is stored and dissipated.

In a steady state, conservation of electrons implies that the rate at which electrons enter the edge region by the reversible uphill flow must be equal to the rate at which they leave the edge region by the irreversible downhill flow. Considering a mid-plateau situation, where the Fermi level lies between the v th and the $v+1$ th Landau level, the conversion mechanism is the following: as a result of the uphill flow, edge states in the v th and lower Landau levels, which for $I=0$ were unoccupied, will now be filled up to such an extent that electrons tunnel (or are inelastically scattered) into states in the almost empty $v+1$ th Landau level at the rate required to obtain a steady state.

2. The downhill flow of excitations

Initially we focus on a region of the sample, the shaded region in Fig. 1, which is far away from source and drain (and other contacts) and which has a width L_y , which is much larger than the magnetic length, but still so small that the Hall field can be treated as constant within the region.

Later we shall discuss the conditions prevailing in the edge regions.

Experiments show that the longitudinal resistivity increases dramatically when the Fermi level lies in the region of the extended states. This shows that scattering processes among extended states are much more efficient than scattering between localized states. At very low temperatures the longitudinal conductivity is determined by scattering between localized states, but at somewhat higher temperatures there is a regime where scattering between extended states dominates entirely [1], and it is this regime we shall discuss in the following. The regions with extended states have small but finite widths. However, we shall use the approximation where, at $I=0$, the extended states of a given Landau band all have the same energy.

The density of excitations (electrons in the $v+1$ th Landau level, and holes in the v th Landau level) is exponentially small in the regime we consider. Blocking of scattering processes because of the Pauli principle can therefore be neglected. The density of electrons in extended states in the $v+1$ th Landau level is

$$n_e = n_{v+1} \cdot f(\epsilon_{v+1}) \simeq n_{v+1} \exp\left(-\frac{\epsilon_{v+1}-\mu}{kT}\right) \quad (1)$$

where n_{v+1} is the density of extended states in the $v+1$ th Landau level and f is the Fermi function. Correspondingly the density of holes in the v th Landau level is

$$n_h \simeq n_v \exp\left(-\frac{\mu-\epsilon_v}{kT}\right) \quad (2)$$

In the current carrying state, because of the Hall field, the Landau levels become tilted such as indicated in Fig. 2,

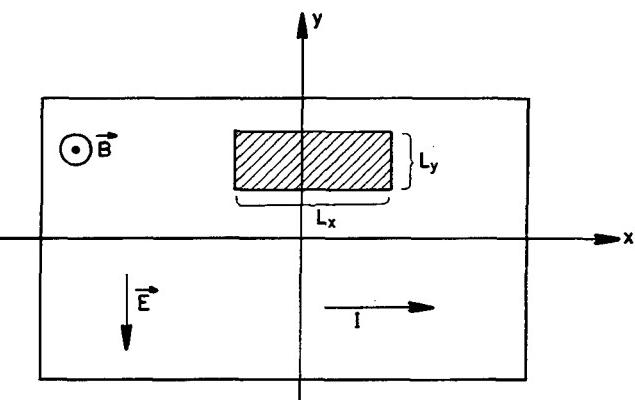


Fig. 1. Geometry of the sample. The current I is flowing in the x -direction. The magnetic field points in the z -direction and the Hall field is pointing in the negative y -direction, i.e. $E_y < 0$.

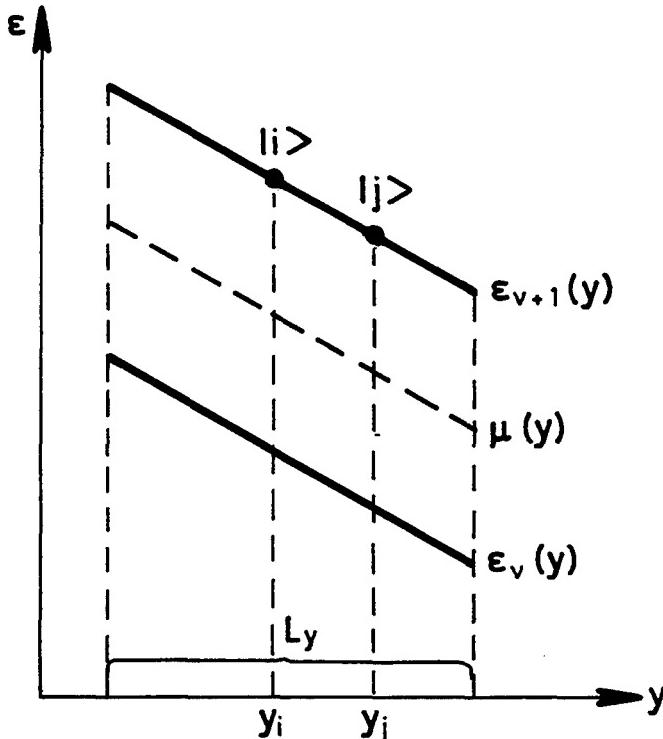


Fig. 2. Landau levels for $I \neq 0$. The Fermi level, $\mu = \mu(y)$, lies between the v th and the $v + 1$ th Landau level.

which depicts the energy of the extended states (of the v th and the $v + 1$ th Landau level) as a function of their centre of mass coordinate $y_i = \langle i | y | i \rangle = \int \Psi_i^* y \Psi_i dx dy$. We consider the scattering of an electron from the state $|i\rangle$ to the state $|j\rangle$ in the $v + 1$ th Landau level. Since the energy ε_j is smaller than ε_i , energy conservation requires the emission of a phonon with the energy $\hbar\omega_q = \varepsilon_i - \varepsilon_j = -\nabla\mu(y_j - y_i) = e|E_y|(y_j - y_i)$. The stated scattering process carries the electron the distance $y_j - y_i$ in the y -direction. Because of disorder, which breaks the lattice-translational invariance, crystal momentum is not conserved. The reversed process, where by the absorption of a phonon, an electron is scattered from the state $|j\rangle$ to the state $|i\rangle$, is less probable than the original process, because whereas the probability for phonon emission is proportional to $(n_q + 1)$, the probability for phonon absorption is proportional only to n_q , where n_q is the average number of phonons of the given kind. This asymmetry between emission and absorption implies that the overall transverse motion (due to inelastic scattering processes of the electrons in the $v + 1$ th Landau level) is downhill, i.e. in the direction of the y -axis. More specifically the probability per second for the process $|i\rangle \rightarrow |j\rangle$ can be written as

$$P_{ij} = K_{ij} f(\varepsilon_i)[1 - f(\varepsilon_j)](n_q + 1) \quad (3)$$

where K_{ij} (except for a constant) is the absolute square of a matrix element, which, via the electron-phonon interaction, connects the two states. And where, according to Fig. 2, $f(\varepsilon_i) = f(\varepsilon_j)$. The probability per second for the process $|j\rangle \rightarrow |i\rangle$ is

$$P_{ji} = K_{ji} f(\varepsilon_j)[1 - f(\varepsilon_i)]n_q \quad (4)$$

with $K_{ij} = K_{ji}$. It is seen that

$$P_{ij} - P_{ji} = K_{ij} f(\varepsilon_i)[1 - f(\varepsilon_i)] \simeq K_{ij} f(\varepsilon_i) \quad (5)$$

showing that the overall contribution, from scattering processes between any two states $|i\rangle$ and $|j\rangle$, to the downhill motion of the electrons, is proportional to the occupational probability and thus to the density of electrons n_e in the $v + 1$ th Landau level. Since the current density of these electrons can be written as $n_e \mu_e E_y e$, we conclude that the transverse mobility μ_e must be independent of the temperature. The same holds for the mobility, μ_h , of the holes in the v th Landau level. Electron-electron interaction does not contribute to the downhill flow because if one electron is scattered a certain distance downhill energy conservation requires that another electron is scattered the same distance uphill. Using (1) and (2) we find the following expression for the transverse excitation current density

$$J_{\text{exc.}} = \left[en_v \exp\left(-\frac{\mu - \varepsilon_v}{kT}\right) \mu_h + en_{v+1} \times \exp\left(-\frac{\varepsilon_{v+1} - \mu}{kT}\right) \mu_e \right] E_y \quad (6)$$

3. The uphill flow of the electron liquid

In a standard Hall bar experiment the total transverse current must be zero. At finite temperatures the system can satisfy this requirement only by establishing a longitudinal electric field E_x . Let us start by discussing the influence of a finite longitudinal field within the free electron model, where the basis consists of wavefunctions each of which is a product of a plane wave in the x -direction and a harmonic oscillator wave function in the y -direction. The centre of mass coordinate, y_0 , of the oscillator wave function is determined by the momentum in the perpendicular direction

$$y_0 = \frac{p_x}{eB} + \frac{|E_y|}{\omega_c B}. \quad (7)$$

However, for $E_x \neq 0$, p_x is no longer a constant of motion. Instead we have

$$\frac{dp_x}{dt} = -eE_x \quad (8)$$

which by use of (7) implies

$$v_y = \frac{dy_0}{dt} = \frac{1}{eB} \frac{dp_x}{dt} = -\frac{E_x}{B} \quad (9)$$

meaning that the entire electron system moves to the left in Fig. 2, i.e. in the negative y -direction. The presence of a finite longitudinal field results in a reversible transverse uphill flow of the electron liquid. The current density associated with this flow is

$$J_{\text{liq.}} = n(-e)v_y = v \frac{eB}{h} e \frac{E_x}{B} = v \frac{e^2}{h} E_x. \quad (10)$$

This, of course, is not very surprising, because the very essence of the QHE, at $T = 0$, is precisely that the electron liquid responds to the presence of an electric field with a current density which is perpendicular to that field [2]. At finite temperatures the electron liquid responds in the same way, but now the scattering of excitations also contributes to the total current density. For $T \neq 0$ the longitudinal field E_x is finite. Unavoidably, therefore, the current density associated with the flow of the electron liquid must have a

component which is perpendicular to the x -axis, and which is given by (10).

As we pass a given transport current I through the system, it responds by generating a Hall field. The presence of the Hall field causes excitations to be scattered downhill, so that a transverse current density $J_{\text{exc.}}$ results. To obtain a steady state the system is bound to establish a longitudinal electric field with such a strength that $J_{\text{liq.}}$ cancels $J_{\text{exc.}}$.

$$J_{\text{liq.}} + J_{\text{exc.}} = v \frac{e^2}{h} E_x + J_{\text{exc.}} = 0 \quad (11)$$

4. Steady state conditions at the edges

In order to make a circulation out of the two opposite flows $J_{\text{exc.}}$ and $J_{\text{liq.}}$, we must consider what happens at the edges. Consider for example the edge that is below in Fig. 1 and to the left in Fig. 2. By the uphill motion of the electron liquid, this edge region receives electrons at the rate $J_{\text{liq.}}/e$ per unit length. To obtain a steady state, electrons must leave the edge region at the same rate. Electrons leave the edge region by scattering and tunnel processes in the v th and the $v+1$ th Landau level. In the v th level it is more convenient to describe the matter in terms of holes, meaning that electrons are scattered out of the edge region at the rate $|J_{\text{exc.}}^{\text{hole}}|/e$ per unit length, where $J_{\text{exc.}}^{\text{hole}}$ is the hole contribution to the excitation current density. To account for the flow of electrons in the $v+1$ th Landau level we must invoke the edge states. Edge states in the v th and lower Landau levels, which for $I=0$ were unoccupied, will now be filled up to such an extent that electrons tunnel (or are inelastically scattered) into states in the almost empty $v+1$ th Landau level at the rate required to obtain a steady state, i.e. at the rate $J_{\text{liq.}}/e - |J_{\text{exc.}}^{\text{hole}}|/e$ per unit length. In this way electrons, which came up the hill by the reversible motion of the electrons in the v th and lower Landau levels, are circulated back (downhill) across the sample via the dissipative electron-phonon processes taking place within the $v+1$ th Landau level. At the other edge the electrons coming downhill tunnel or are scattered into empty states of the v th and lower Landau levels.

5. Energy conversions

Let us start by considering ordinary conduction in a wire. The rate of work done by external sources is $V_L \cdot I$, where V_L is the voltage drop along the wire. This work is initially stored in the form of kinetic energy of the electrons. However, in a steady state, because of friction, this kinetic energy is dissipated at the same rate as it is stored. Then consider the quantum Hall effect. In this case the rate of work done by external sources is

$$V_L \cdot I = L_x E_x \cdot L_y v \frac{e^2}{h} |E_y| \quad (12)$$

The reversible uphill flow of the electron liquid stores potential energy at the rate

$$J_{\text{liq.}} \cdot |E_y| = v \frac{e^2}{h} E_x |E_y| \quad (13)$$

per unit area. This rate is equal to the rate of work per unit area $V_L \cdot I / L_x L_y$, showing that all the work is initially stored as potential energy. In a steady state the rate (13) is

equal to the rate of dissipation per unit area $J_{\text{exc.}} E_y = \sigma_{xx} E_y^2$. The role played by kinetic energy in ordinary conduction, in the quantum Hall effect is played by the potential energy.

6. The temperature independence $\Delta\rho_{xy}/\rho_{xx}^{\min}$

In terms of conductivities the requirement that the total transverse current must be zero takes the form $0 = \sigma_{yx} E_x + \sigma_{xx} E_y$. From (11) and (6) it follows that σ_{xx} is equal to the expression in the bracket on the right side of equation (6). Suppose we can vary μ , say by changing the electron density. We can then determine the value of μ that minimizes σ_{xx} . We find

$$\frac{\partial \sigma_{xx}}{\partial \mu} = 0 \rightarrow \exp \frac{\mu}{kT} = \left(\frac{n_v \mu_h}{n_{v+1} \mu_e} \right)^{1/2} \exp \left(\frac{\epsilon_v + \epsilon_{v+1}}{2kT} \right) \quad (14)$$

implying

$$\sigma_{xx}^{\min} = 2e(n_v \cdot n_{v+1} \cdot \mu_e \cdot \mu_h)^{1/2} \exp \left(- \frac{\epsilon_{v+1} - \epsilon_v}{2kT} \right) \quad (15)$$

We now want to determine $\Delta\sigma_{yx} = \sigma_{yx}(T) - \sigma_{yx}(0)$ at the value of μ , where σ_{xx} is minimal. In the free electron model a filled Landau level gives rise to a current density $J_x = e^2/h (-E_y)$. If all the extended states in a Landau band are occupied, again, because of the compensation theorem [3], we have the same current density. Therefore, if the extended states of the $v+1$ th Landau level were all filled, they would contribute to the current density by the amount $e^2/h (-E_y)$. However, since only the fraction $\exp(-(\epsilon_{v+1} - \mu/kT))$ of these states is occupied we get the contribution $e^2/h \exp(-(\epsilon_{v+1} - \mu/kT)) (-E_y)$. Similarly, the presence of holes in the v th Landau level reduces the current density by the amount $e^2/h \exp(-(\mu - \epsilon_v/kT)) (-E_y)$. Thus it follows that the current density at the finite temperature T is changed relative to the zero temperature case by the amount

$$\Delta J_x = \frac{e^2}{h} \left[\exp \left(- \frac{\epsilon_{v+1} - \mu}{kT} \right) - \exp \left(- \frac{\mu - \epsilon_v}{kT} \right) \right] (-E_y) \quad (16)$$

We have $\Delta J_x = \Delta\sigma_{yx} (-E_y)$, where we have used $-\sigma_{xy} = \sigma_{yx}$. From (14) and (16) it then follows that

$$\Delta\sigma_{yx} = \frac{e^2}{h} \left[\left(\frac{n_v \mu_h}{n_{v+1} \mu_e} \right)^{1/2} - \left(\frac{n_{v+1} \cdot \mu_e}{n_v \mu_h} \right)^{1/2} \right] \times \exp \left(- \frac{\epsilon_{v+1} - \epsilon_v}{2kT} \right) \quad (17)$$

Using that

$$\frac{\Delta\sigma_{xy}}{\rho_{xx}^{\min}} = - \frac{\Delta\sigma_{xy}}{\sigma_{xx}^{\min}}$$

we find

$$\frac{\Delta\sigma_{xy}}{\rho_{xx}^{\min}} = \frac{e}{2h} \left[\frac{1}{n_v \cdot \mu_h} - \frac{1}{n_{v+1} \cdot \mu_e} \right] \quad (18)$$

which is temperature independent because the mobilities are temperature independent.

7. Discussion and conclusions

The QHE displays several interesting non-local effects [4]. One might therefore question the meaningfulness of applying a local relationship between the electric field and the current density. However, in a recent investigation [5], it has been argued that a considerable part of the transport current flows in the bulk, and that here there exists a quasi-local relationship between the electric field and the current density, when the Fermi level lies in a region with localized states. In the valleys and around the mountains of the disorder potential, we have circulating currents, which are completely unrelated to the macroscopic electric field, induced by the transport current. Let J_0 denote the current density associated with these circulating currents for the $I = 0$ case. For $I \neq 0$, we can then define the excess current density $J = J_{\text{actual}} - J_0$, and it is this excess current density, which, in the bulk, as a reasonable approximation, can be linearly related to the macroscopic electric field at the same place.

Although the values of ρ_{xx}^{\min} and $\Delta\rho_{xy}$ increase rapidly with temperature, it has been found experimentally [6] that their ratio is independent of temperature down to the

0.01 ppm level of accuracy. Only an argument which, essentially, involves no approximations can explain such an accuracy. Therefore, the present theory is unable to explain this staggering accuracy.

The picture of a transverse circulation of electrons, superimposed on the longitudinal transport current, provides a physical interpretation of the experimentally imposed condition $J_y = \sigma_{yx} E_x + \sigma_{xx} E_y = 0$ by the identifications: $J_{\text{exc.}} = \sigma_{xx} E_y$ and $J_{\text{liq.}} = \sigma_{yx} E_x$, and by explaining how the two flows are converted into each other at the edges. Transverse circulation of electrons also applies to the FQHE, except that in this case the excitation current is carried by quasi-electrons and quasi-holes.

References

1. Das Sarma, S. and Dongzi Liu, Phys. Rev. B, **48**, 9166 (1993).
2. Dolgopolov, V. T. et al., Phys. Rev. B, **46**, 1250 (1992).
3. Prange, R. E., Phys. Rev. B, **23**, 4802 (1981).
4. McEuen, P. L. et al., Phys. Rev. Lett. **64**, 2062 (1990).
5. Thouless, D. J., Phys. Rev. Lett. **71**, 1879 (1993).
6. Cage, M. E. et al., Phys. Rev. B, **30**, 2286 (1984).

Spin Resonance Determination of the Effective g-Factor of Electrons in Low Dimensional (GaIn)As/InP Structures

B. Kowalski and P. Omling

Department of Solid State Physics, University of Lund, Box 118, S-221 00 Lund, Sweden

B. K. Meyer and D. M. Hofmann

Physikdepartment E16, Technical University of Munich, D-85747 Garching, Germany

V. Härle and F. Scholz

4. Physikalisches Institut, University of Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany

and

P. Sobkowicz

Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668 Warsaw, Poland

Received April 28, 1994; accepted June 15, 1994

Abstract

Optically detected magnetic resonance measurements of the effective g -value (g^*) of electrons in type-I $\text{Ga}_x\text{In}_{1-x}\text{As}/\text{InP}$ quantum wells are presented. The observation of spin resonance on the circularly polarized luminescence is explained in terms of recombination in a one-side p-modulation doped quantum well and spin thermalization. Quantum confinement changes the electron effective g -value. It further induces a strongly anisotropic g -tensor. A calculation for the $g_{||}^*$ -component agrees well with the experimental data, but the anisotropy can be explained only qualitatively.

Introduction

Spin-orbit coupling in semiconductors leads within $\mathbf{k} \cdot \mathbf{p}$ -theory to the definition of the effective g -factor g^* [1, 2]. The motivation for this is very much the same as that for the definition of the effective mass m^* . As m^* describes the influence of the crystal potential on the space coordinates of an electron, and their time derivatives, g^* represents the effect of the crystal potential on the spin coordinates. Equivalently to the effective mass it describes the semiconductor's band structure. The band structure is well known for most bulk semiconductors, but it is less thoroughly understood in low-dimensional structures (quantum wells, quantum wires, quantum dots). This understanding can be obtained by measuring the band structure parameters m^* and/or g^* , which are then compared with theory, e.g. $\mathbf{k} \cdot \mathbf{p}$ -calculations.

It is well known that magnetic resonance techniques, such as electron spin resonance (ESR) and cyclotron resonance (CR), allow the most precise determination of g^* and m^* . However for reduced dimensionality several obstacles are encountered. First, the reduction of sample volume increases the demands on sensitivity making optical detection necessary (optically detected ESR (ODESR) [3] and optically detected CR (ODCR) [4]). Second, scattering increases with decreasing size of the low dimensional structure, thus rendering the measurement more imprecise. Here ESR seems to be less affected than CR [5]. Furthermore the role of CR can be questioned, when the size of, e.g., a quantum dot is smaller than the cyclotron radius.

The main drawback of ODESR, however, originates from the prerequisite for observing the spin resonance: it is the introduction of spin transitions within the lifetime of the radiatively recombining carriers. For the standard ODESR technique using cw microwaves the time needed for inducing a spin transition is experimentally limited to be larger than 0.1 μs . Only a few experimental results of ODESR are existing, which have been obtained for type-II superlattices [6–9]. Here the electrons and holes are separated in real space. This separation of the carriers leads to optical life times in the microsecond range. However, in the technologically important type-I (direct in space) QW systems, like $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{GaAs}$ or $\text{InP}/\text{In}_{1-x}\text{Ga}_x\text{As}$, such investigations have in general failed, since the radiative recombination times are too small.

In type-I QW's longer lifetimes can be obtained in samples to which an electric field perpendicular to the QW plane was applied by means of electrical contacts [10]. This is, however, incompatible with the ODESR technique using a microwave cavity. We used instead a single-sided, p-modulation doping for creating an electric field and thus reducing the radiative recombination rates. The data for electron g^* -values in $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ QWs show a clear dependence of g^* on QW width and a strong, confinement-induced anisotropy of the g -tensor. In this article we focus on how circular polarization of the photoluminescence (PL) is connected to the spin states and how the magnetic resonance signal is detected.

Experimental

The $\text{Ga}_x\text{In}_{1-x}\text{As}/\text{InP}$ single QW samples, or samples containing several QWs of different thicknesses, used in this investigation have been grown by low pressure metal organic vapour phase epitaxy at 620 °C on semi-insulating InP : Fe as described elsewhere [11]. The QWs were lattice matched ($x_{\text{Ga}} = 0.47$) with varying QW thickness (d). The structure is the same for all samples: the QW is grown on top of a 400 nm InP buffer layer followed by a 5 nm spacer and a

single-sided p-modulation doping, a 10 nm layer with an acceptor concentration of $1-2 \times 10^{18} \text{ cm}^{-3}$, and finally a 60 nm capping layer.

The PL was excited by an Ar⁺ laser operating at 514 nm with typical incident power densities 0.1–0.3 W/cm². The PL was dispersed with a f/0.22 m double grating spectrometer and detected with a Ge photodiode cooled to 77 K. The degree of circular polarization, i.e. the difference of right-(σ^+) and left-circularly (σ^-) polarized components of the PL, ($\sigma^+ - \sigma^-$) was analysed by a photoelastic modulator and a linear polarizer placed in front of the monochromator. The samples were immersed in pumped LHe at ~ 1.6 K. They were located in an open TE₀₁₁ cylindrical microwave cavity, itself placed in the center of a 4T split coil superconducting magnet. The ODES experiments were carried out by applying 24 or 36 GHz microwaves, scanning the magnetic field and recording ($\sigma^+ - \sigma^-$) of distinct PL features in Faraday configuration using the monochromator as band pass filter with 2–8 meV band width.

Results

The PL spectra from all the Ga_{0.47}In_{0.53}As/InP QWs were determined by single peaks at energies between 0.83 eV ($d = 15$ nm) and 0.89 eV ($d = 6$ nm). Full peak width at half maximum were between 8 meV and 15 meV respectively. These PL features are due to direct recombination of free carriers between the first electron and heavy hole subbands. Excitons are quenched because of the two-dimensional hole gas present in the QWs [12]. Due to the one side modulation doping induced electric field the energies of the PL peaks are slightly shifted away from the values usually observed in symmetrical QWs of this material system.

When applying a magnetic field a circular polarization of the PL was observed. The degree ($\sigma^+ - \sigma^-$) of this circular polarization was linearly dependent on magnetic field for fields below 2 T. It was below 4% at a field of 1 T. When measuring ($\sigma^+ - \sigma^-$) as a function of magnetic field at 1.6 and 4.2 K, the slope at 4.2 K was typically 0.7 times the one at 1.6 K.

Switching the microwaves on, the ODES was observed as a resonant decrease in the ($\sigma^+ - \sigma^-$) signal at a certain field. This resonant magnetic field changed when rotating the sample through an angle θ from $B \parallel \langle 001 \rangle$, i.e. parallel to the QW axis with θ being zero, towards $B \parallel \langle 110 \rangle$, i.e. lying in the QW plane. As an example the ODES spectra of a 10 nm wide Ga_{0.47}In_{0.53}As/InP QW are shown in Fig. 1 for four angles θ . The maximum angle of detection was between 70° and 80°, since the projection of ($\sigma^+ - \sigma^-$) along the direction of detection decreased with increasing angle. Here ($\sigma^+ - \sigma^-$) is defined with respect to the QW axis.

Using the resonant magnetic field B_{res} and the constant microwave energy $\Delta E_{\mu\omega}$ the g-values were calculated in spin $S = \frac{1}{2}$ formalism by the relation $\Delta E_{\mu\omega} = g^* \mu_B B_{\text{res}}$ (μ_B is the Bohr magneton). The g-values as a function of θ were analysed using the standard expression for a g-tensor in axial symmetry:

$$g^*(\theta) = \sqrt{g_{||}^*{}^2 \cos^2 \theta + g_{\perp}^*{}^2 \sin^2 \theta}. \quad (1)$$

A fit to the experimental g-values for the 10 nm wide Ga_{0.47}In_{0.53}As/InP QW is shown in the inset of Fig. 1, where g_{\perp}^* has been used as the free parameter. In this

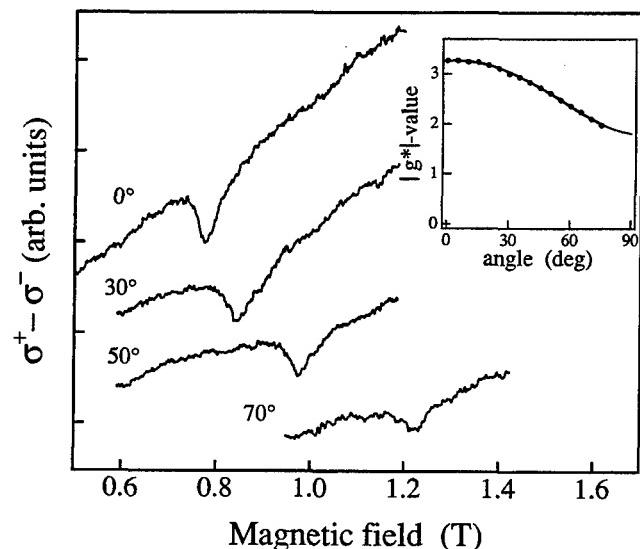


Fig. 1. ODMR (36 GHz) detected on the magnetic circularly polarized PL of a 10 nm Ga_{0.47}In_{0.53}As/InP SQW at $T = 1.6$ K for four different angles. The inset shows the anisotropy of g^* . Solid circles are experimental data and the solid line is a fit using the expression for axial symmetry (see text).

manner values for $g_{||}^*$ and g_{\perp}^* were obtained for lattice matched QWs with thicknesses between 6 and 15 nm, as well as for a quasi three-dimensional $d = 100$ nm thick reference layer. The results are summarized in Fig. 2.

With varying quantum confinement the absolute values of $g_{||}^*$ and g_{\perp}^* , as well as the ratio $g_{||}^*/g_{\perp}^*$ changed. In the quasi three-dimensional case a $g^* = 4.01 \pm 0.04$ resonance was obtained. The resonance was isotropic within the experimental uncertainty. With increasing quantization the anisotropy increases to $g_{||}^*/g_{\perp}^* = 4$ for $d = 6$ nm, the thinnest QW for which ODES could be observed.

Discussion

For an explanation of the experimental findings we first look at the band structure of the (InGa)As/InP QWs. Quantum confinement lifts the four fold degeneracy of the valence band edge pushing the first heavy hole states $|J, M_J\rangle = |\frac{3}{2}, \pm \frac{3}{2}\rangle$ above the highest light hole state $|J, M_J\rangle = |\frac{3}{2}, \pm \frac{1}{2}\rangle$. The σ^+ (σ^-) component of the PL originates from transitions from the $|S, M_S\rangle = |\frac{1}{2}, -\frac{1}{2}\rangle$ ($|\frac{1}{2}, +\frac{1}{2}\rangle$) electron state to the $M_J = -\frac{3}{2}$ (+ $\frac{3}{2}$) heavy hole state

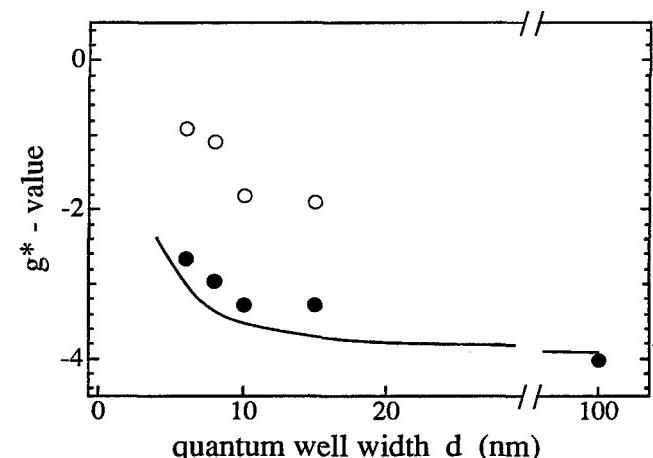


Fig. 2. $g_{||}^*$ (solid circles) and g_{\perp}^* (open circles) for different well widths in Ga_{0.47}In_{0.53}As/InP QWs. The solid line is the result of a calculation (see text).

[13] (see Fig. 3). The heavy hole is the only occupied hole state. In case of equal population in the $M_s = \pm \frac{3}{2}$ and $M_s = \pm \frac{1}{2}$ Kramers doublets the PL does not show circular polarization, since the σ^+ and σ^- components have equal transition probability. With a magnetic field electrons and holes will thermalize into the lowest of the Zeeman split states respectively, and are hence polarized. However, the ratio of the hole- (P_h) and electron-polarization (P_e) is determined by $P_h/P_e = \text{const} \times g_h 2kT/g_e E_F$ [14], where kT , g_h and g_e are the thermal energy, hole- and electron g -value respectively. We estimate the Fermi energy E_F for the two dimensional hole gas ($N_s \approx 10^{12} \text{ cm}^{-2}$) to be 5–20 meV. It is therefore certainly larger than $kT \approx 0.14 \text{ meV}$. Thus the hole polarization can be neglected, and only the population difference within the electron Kramers doublet gives rise to circular polarization of the PL.

This population difference is a necessary condition for performing spin resonance. However, the photo generated electrons at the conduction band edge seem to thermalize only partly, before recombining radiatively. In case of equilibrium the $(\sigma^+ - \sigma^-)$ -signal would at low magnetic fields be proportional to T^{-1} [14]. This was not found in our experiments. Under magnetic resonance conditions the induced spin transitions transfer population from the $M_s = -\frac{1}{2}$ to the $M_s = +\frac{1}{2}$ states and hence alter the $(\sigma^+ - \sigma^-)$ -signal. Here the $M_s = -\frac{1}{2}$ state is lowest in energy, since σ^+ and σ^- are connected to $M_s = -\frac{1}{2}$ and $+\frac{1}{2}$ states respectively, and $(\sigma^+ - \sigma^-) > 0$. From the resonant magnetic field the g -value is calculated. Its sign can also be derived: The sign of the magnetic field with respect to detection direction, in our case negative, leads to negative g -values.

The observed anisotropy in the spin resonance signal could, in accordance with previous studies of type-II systems, be taken as evidence for a hole spin resonance [7, 8]. For heavy holes an analysis for the $|J, M_J\rangle = |\frac{3}{2}, \pm \frac{3}{2}\rangle$ states gives $g_{\parallel}^* = 4$ and $g_{\perp}^* = 0$ [7, 9]. The discrepancy between the present results and these values is therefore a strong indication that holes are not involved. Furthermore the Fermi level of the holes is located several meV below the valence edge. A hole-resonance (at $k = k_F$) is highly unlikely since the magnetic resonance is detected on the band edge PL ($k = 0$). Here both hole spin states are occupied rendering spin resonance impossible.

The observation of ODESR signals in type-I QW structures is apparently facilitated by the spatial separation of the electron and hole wave-functions caused by the single-sided modulation doping. Self-consistent calculations of overlap integrals give an increase of approximately two orders of magnitude, corresponding to an increase of the optical lifetime from the intrinsic $\approx 1 \text{ ns}$ to $\approx 0.1 \mu\text{s}$, the region where spin-flip transitions can be made to occur. Another possible reason may be the admixture of p -type states in the pure spin-up and spin-down s -type wavefunctions of the conduction band due to the electric field, $\mathbf{k} \cdot \mathbf{p}$ interaction and inversion asymmetry [15]. This could, in principle, lead to a relaxation of the pure magnetic dipole selection rules and an increased spin-flip probability through electric dipole spin transitions. Time-resolved measurements to further elucidate this question are in progress.

The experimental technique used in this investigation is based on an occupation difference in the spin population. It is, however, well known that the relaxation of the electronic spin polarization is accompanied by a build up of a nuclear spin polarization via flip-flop processes (Overhauser effect) [16]. The result is that the polarized nuclei produce a magnetic field that introduces a shift of the spin resonance position. An estimate [17], including Ga-, As- and In-nuclei, gives in our case a 10 mT shift, being within the experimental uncertainty of g^* .

Bulk $\mathbf{k} \cdot \mathbf{p}$ theory can explain qualitatively the dependence of the experimental g^* -values on QW width, as well as the large anisotropy of the g -tensor. In this first, quasi three-dimensional approach interface- and asymmetric potential effects on the spin splitting [18] are neglected. g_{\parallel}^* reads:

$$\frac{g_{\parallel}^*}{g_0} = 1 - \frac{2}{3m} P^2 \left(\frac{3}{2(E_g + E_{e1} - E_{hh1})} - \frac{1}{2(E_g + E_{e1} - E_{lh1})} - \frac{1}{(E_g + E_{e1} - E_{sh})} \right), \quad (2)$$

and g_{\perp}^* :

$$\frac{g_{\perp}^*}{g_0} = 1 - \frac{2}{3m} P^2 \left(\frac{1}{(E_g + E_{e1} - E_{lh1})} - \frac{1}{(E_g + E_{e1} - E_{sh})} \right), \quad (3)$$

where in the g^* -values for non-degenerate heavy- and light-hole bands [2] the bandgap energy E_g , the valence band spin orbit splitting E_{sh} and two dimensional subband energies have been used. The subband energies E_{e1} , E_{hh1} and E_{lh1} for the lowest electron, heavy- and light-hole states respectively were determined for a flat QW potential. Equations (2) and (3) explain the increase of g_{\parallel}^* and g_{\perp}^* towards zero. The difference of equations (2) and (3) describes the anisotropy of the g -tensor:

$$\frac{g_{\parallel}^* - g_{\perp}^*}{g_0} = \frac{2}{3m} P^2 \left(\frac{1}{(E_g + E_{e1} - E_{hh1})} - \frac{1}{(E_g + E_{e1} - E_{lh1})} \right) \quad (4)$$

It increases with decreasing QW width. However, the calculated g^* -values deviate by up to 30% from the experimental values. A calculation, including either higher order terms in \mathbf{k} or using a Kane-model, was done by Ivchenko *et al.* [19] for the GaAs/(AlGa)As system. Adapted to $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ it can lead to a better fit. It still does not include interface- and asymmetric potential effects on the spin splitting.

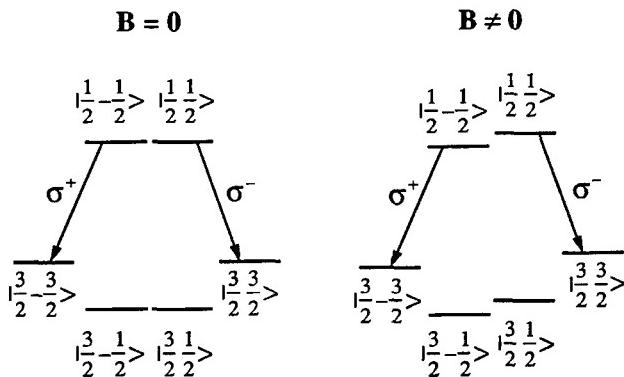


Fig. 3. The valence- and conduction band edges in spin representation with and without magnetic field. The circular components σ^+ and σ^- are the radiative transitions relevant in our experiments.

Therefore we used a more realistic model for calculating the g_{\parallel}^* -values as a function of QW width in the $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ system [20]. It leads to quantitatively good agreement with the experimental data (see Fig. 2). In this self-consistent, sub-band calculation an 8×8 Kane-Hamiltonian is used. In the calculation of the spin-splitting, the Zeeman interaction, the effects of the asymmetric potential, and the interface effects, are taken into account. The material parameters are taken from Ref. [21], and the bulk Zeeman splitting factor $g^* = -4.1$ was chosen to fit the experimental data. The agreement with experiment is quite satisfactory, bearing in mind that the inversion-asymmetry spin-splitting, proportional to k^3 , has not been taken into account. It is, however, a difficult task to calculate the g_{\perp}^* -value, since an accurate choice of basis-functions is non-trivial.

The contributions of asymmetric potential- and interface effects to the spin splitting deserve a closer look: First, in an asymmetric QW potential the energy levels differ from the values for a flat QW. This introduces corrections of the g^* -values. This can be viewed as an indirect effect. Second, the asymmetric potential and the $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ interfaces directly induce a spin splitting [18]. This we will refer to as electrically induced spin splitting. Both contributions cause also the g^* -value in the 100 nm QW to deviate from the true $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}$ bulk value. Therefore we could not directly use this experimental value as a parameter in our calculations. Instead we found $g^* = -4.1$ to be a reasonable bulk value. This is in agreement with the value suggested in ref. [22] using electrically detected spin resonance in a 2DEG, but is quite different from $|g^*| = 5.2$ value reported in Ref. [23].

In our calculations the electrically induced spin splitting renders the g^* -values to be strongly dependent on magnetic field at low fields. Since our measurements at 24 GHz and 36 GHz gave identical results we conclude that we are safely above this magnetic field region. The electrically induced spin splitting could, in fact, be the reason for the weak $|g^*| = 5.6$ signal observed in a lattice-matched 15 nm QW [24], where low microwave frequencies have been used. This is supported by an independent investigation on exactly the same sample in which an electron-hole separation in the QW was reported [25]. Another reason for their result may be that they measured on a bound electron state.

In conclusion, we have measured the conduction band spin splitting in type-I $\text{Ga}_{0.47}\text{In}_{0.53}\text{As}/\text{InP}$ QWs. This was possible due to a one sided p-modulation doping, which furthermore related the circular polarization of the PL to the electrons only. The mechanisms for observing spin resonance are explained in detail. The clear dependence of the g^*

on quantum confinement has been described theoretically. Quantum confinement further induced a strong anisotropy of the g -tensor. However, g_{\perp}^* could not be calculated sufficiently exact.

The technical support of H. Bolay and E. Lux is acknowledged. This work has been financed by the Swedish Natural Science Research Council and the Deutsche Forschungsgemeinschaft, B.K. is grateful for a scholarship from the Friedrich Ebert Foundation.

References

1. Roth, L. M., Lax, B. and Zwerdling, S., Phys. Rev. **114** 90 (1959).
2. Hendorfer, G. and Schneider, J., Semicond. Sci. Technol., **6**, 595 (1991).
3. Herman, C. and Weisbuch, C., Phys. Rev. **B15**, 816 (1977).
4. Baranov, P. G., Veshchunov, Y. P., Zhitnikov, R. A., Romanov, R. A. and Shreter, Y. G., Pis'ma Zh. Eksp. Teor. Fiz. **26**, 369 (1977) [JETP Lett. **26**, 249 (1977)].
5. Dobers, M. et al., "High Magnetic Fields in Semiconductor Physics II" (Edited by G. Landwehr) (Springer, Berlin 1989), p. 386.
6. van Kesteren, H. W. et al., Phys. Rev. Lett. **61**, 129 (1988).
7. Glaser, E. et al., Phys. Rev. Lett. **65**, 1247 (1990).
8. Trombetta, J. M., Kennedy, T. A., Gammon, D. and Prokes, S. M., in "20th Int. Conf. on Semicond. Phys." (Edited by E. M. Anastassakis and J. M. Joannopoulos) (World Scientific, Singapore 1990), p. 1361.
9. Glaser, E. R. et al., Phys. Rev. **B47**, 1305 (1993).
10. Polland, H. J., Schultheis, L., Kuhl, J., Göbel, E. and Tu, C., Phys. Rev. Lett. **55**, 2610 (1985).
11. Härlé, V. et al., in "Proceedings 5th Int. Conf. on InP and Related Materials" (Paris, France 1993), p. 191.
12. Miller, R. C. and Kleinman, D. A., J. Lum. **30**, 520 (1985).
13. Bastard, G. "Wave Mechanics Applied to Semiconductor Heterostructures" (Les Edition De Physique, JOUVE, Paris 1988).
14. Dyakonov, M. I. and Perel, V. I., in "Optical Orientation" (Edited by F. Meier, B. P. Zakharchrenya) (Elsevier Science Publ., North Holland, Amsterdam 1984).
15. Brown, M. and Rössler, U., J. Phys. C: Solid State Phys. **18**, 3365 (1985).
16. Herman, C. and Weisbuch, C., Phys. Rev. **B15**, 816 (1977).
17. Gotschy, B., Denninger, G., Obloh, H., Wilkering, W. and Schneider, J., Sol. State Commun. **71**, 629 (1989).
18. Sobkowicz, P., J. of Crystal Growth, **101**, 337 (1990).
19. Ivchenko, E. L. and Kiselev, A. A., Sov. Phys. Semicond. **26**, 827 (1992) [Fiz. Tekh. Poluprovodn **26**, 1471 (1992)].
20. Kowalski, B. et al., Phys. Rev. **B49**, 14786 (1994).
21. Landolt-Börnstein, Semiconductors, (Edited by O. Madelung, M. Schulz and H. Weiss,) (Springer Verlag, Berlin 1982), Vol. 17a.
22. Dobers, M. et al., Phys. Rev. **B40**, 8075 (1989).
23. Kana-ah, A., Johnson, G. R., Cavenett, B. C., Skolnick, M. S. and Bass, S. J., Phys. Rev. **B36**, 1303 (1987).
24. Johnson, G. R., Kana-ah, A., Cavenett, B. C., Skolnick, M. S. and Bass, S. J., Semicond. Sci. Technol. **2**, 182 (1987).
25. Anderson, D. A., Bass, S. J., Kane, M. J. and Taylor, L. L., Appl. Phys. Lett. **49**, 1360 (1986).

Characteristic Potentials for Mesoscopic Rings Threaded by an Aharonov–Bohm Flux

M. Büttiker

IBM T. J. Watson Research Center, Yorktown Heights, N.Y. 10598

Received May 4, 1994; accepted June 14, 1994

Abstract

Electro-static potentials for samples with the topology of a ring and penetrated by an Aharonov–Bohm flux are discussed. The sensitivity of the electron-density distribution to small variations in the flux generates an effective electro-static potential which is itself a periodic function of flux. We investigate a simple model in which the flux sensitive potential leads to a persistent current which is enhanced compared to that of a loop of non-interacting electrons. For sample geometries with contacts the sensitivity of the electro-static potential to flux leads to a flux-induced capacitance. This capacitance gives the variation in charge due to an increment in flux. The flux-induced capacitance is contrasted with the electro-chemical capacitance which gives the variation in charge due to an increment in an electro-chemical potential. The discussion is formulated in terms of characteristic functions which give the variation of the electro-static potential in the interior of the conductor due to an increment in the external control parameters (flux, electro-chemical potentials).

1. Introduction

In this work we discuss the electro-static potential in the interior of small mesoscopic conductors with the topology of a ring (see Fig. 1). An Aharonov–Bohm flux (AB-flux) penetrates the hole of the loop. Typically the electron-density distribution in such a loop is non-uniform and is a sensitive function of the AB-flux. The nonuniform density generates an electro-static potential which is also a function of the flux. We introduce a characteristic potential $v(r)$ which gives the variation of the electro-static potential in the interior of conductors in response to an increment of the

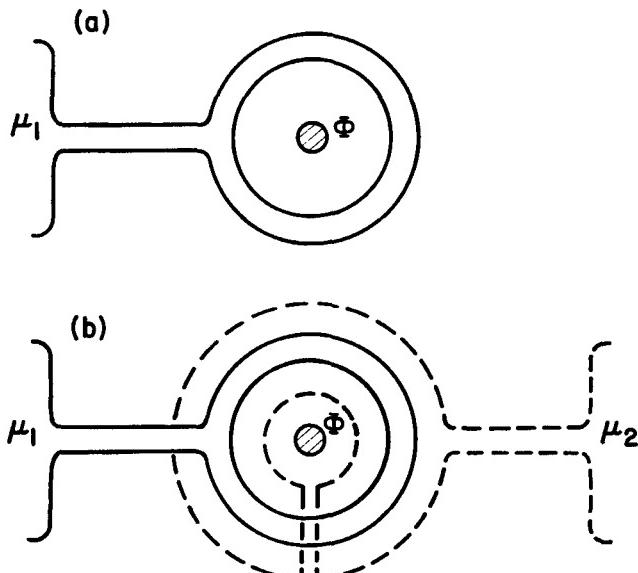


Fig. 1. (a) Ring with a lead connected to an electron reservoir with an electro-chemical potential μ_1 . (b) The ring is separated by an insulator from a second conductor (gate) which is connected to an electron-reservoir at an electro-chemical potential μ_2 .

AB-flux. We investigate a simple example of a one-dimensional loop structure (see Fig. 2) which permits a solution for the characteristic potential $v(r)$. This example provides a demonstration that interaction can enhance the persistent current above its value in the absence of interactions. The characteristic potential $v(r)$ is used to define a flux-induced capacitance which is the ratio of the charge increment on the conductor divided by the increment in flux. The flux-induced capacitance is an odd (periodic) function of flux. For samples with contacts we introduce a characteristic potential $u(r)$ which gives the variation of the local potential inside the conductor in response to an increment of an electro-chemical potential at a contact. This characteristic function allows the evaluation of the electro-chemical capacitance. The electro-chemical capacitance is an even periodic function of flux with period $\Phi_0 = hc/e$. The electro-chemical capacitance can be measured with the help of small time-dependent, oscillating voltages applied to a capacitor. The flux-induced capacitance is measured by applying a small oscillatory flux superimposed on a steady state AB-flux.

Recent work [1] by Thomas, Prêtre and this author on the admittance of mesoscopic capacitors emphasized the need to distinguish between electric and electro-chemical capacitances. Capacitance coefficients are not electro-static entities determined by the geometry of the sample alone but are electro-chemical quantities which depend on the properties of the conductor [2]. The non-geometrical contributions to the capacitance coefficients arise due to the fact that electric fields impinging on a conductor are not screened immediately at the surface of the conductor but penetrate over a distance of a screening length into the “bulk” of the conductor [3]. Clearly field penetration is very important for mesoscopic conductors [1] since one or more dimensions of such

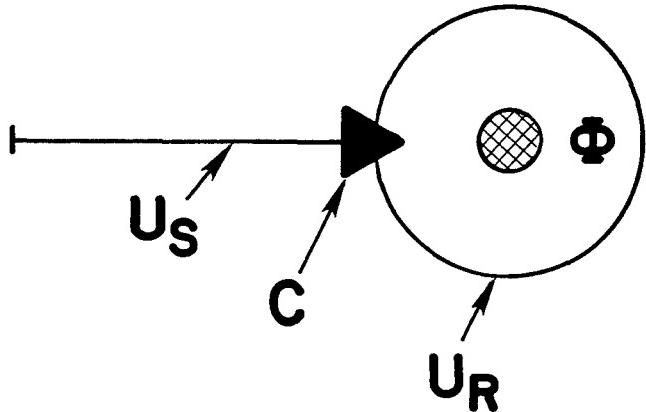


Fig. 2. Normal electron ring which is via a tunneling barrier with capacitance C connected to a short wire (stub) of finite length.

a small system might in fact be comparable to a screening length! A dramatic demonstration of the non-geometrical nature of capacitances is given by Chen *et al.* [4] who show that magnetic fields can completely quench certain elements of a capacitance tensor.

In samples without transmission the capacitance coefficients are an equilibrium phenomenon which reflects properties of the ground state. Another interesting ground state property of a small mesoscopic ring is the equilibrium persistent current. The existence of such currents in mesoscopic disordered normal rings was predicted by this author in collaboration with Imry and Landauer [5]. In earlier work persistent currents were discussed for large molecules [6, 7], Landau diamagnetism [8], and in connection with superconductivity [9, 10]. For many years equilibrium currents in mesoscopic conductors remained of interest to only a small community [11–15]. Experimental observation of such equilibrium currents in an ensemble of rings by Levy *et al.* [16], in single metallic loops by Chandrasekhar *et al.* [17], and single semiconductor rings by Mailly *et al.* [18] generated a considerable theoretical literature. We can refer the reader here to only a few representative contributions investigating the persistent current of non-interacting [19–21] and interacting [22–26] electrons. Persistent currents occur not only in isolated rings but also in rings connected via leads to electron reservoirs [12, 14, 21, 27, 28]. The experiment of Mailly *et al.* [18] in fact measures persistent currents in both closed and open rings of roughly the same amplitude. Since both capacitances and persistent currents reflect properties of the ground state it is intriguing to explore possible connections between the two phenomena.

2. Self-consistent potential of a normal loop

Consider the conductors in Fig. 1. The conductor of Fig. 1(a) is via a lead connected to an electron reservoir [12, 27, 28]. It is otherwise in an electrically isolating environment: There are no electric field lines penetrating the surface of a volume which is at a sufficient distance from the loop and intersects the reservoir far enough from the connection to the lead. Typically conductors are not in such an electrically isolated environment but couple via long range Coulomb forces to other nearby metallic conductors. Such a situation is investigated for the conductor of Fig. 1(b). Here the conductor of Fig. 1(a) is separated via an insulator from a second conductor, a gate, which is connected to an electron-reservoir at electro-chemical potential μ_2 . Again we will for clarity assume that there exists a volume which encloses both conductors and a portion of the reservoirs such that no electric-field lines penetrate its surface. In both structures one of the conductors forms a loop that is threaded by an AB-flux. The equilibrium electro-static potential $U(\mu_1, \mu_2, \Phi, \mathbf{r})$ for these conductors is a function of the electro-chemical potentials, the flux and a complicated function of position. The flux-dependence of the potential is generated by the electron-density of the loop which is flux dependent whenever the loop is not rotationally invariant. We are interested in the variation of the electro-static potential under small changes of the external control parameters μ_1 , μ_2 and Φ . Small increases $d\mu_1$, $d\mu_2$, $d\Phi$ in the electro-chemical potentials and or the flux will bring the conductor from equilibrium state with potential $U(\mu_1, \mu_2, \Phi, \mathbf{r})$ to a

new equilibrium state with potential $U(\mu_1 + d\mu_1, \mu_2 + d\mu_2, \Phi + d\Phi, \mathbf{r})$. The difference dU between these two equilibrium potentials can be expanded in powers of the increment in electro-chemical potential and the increment in flux. To linear order we have

$$e dU(\mu_1, \mu_2, \Phi, \mathbf{r}) = u_1(\mathbf{r}) d\mu_1 + u_2(\mathbf{r}) d\mu_2 + v(\mathbf{r}) d\Phi. \quad (1)$$

Here $u_k(\mathbf{r}) = e dU(\mu_1, \mu_2, \Phi, \mathbf{r})/d\mu_k|_{d\mu_k=0}$, with $k = 1, 2$ and $v(\mathbf{r}) = e dU(\mu_1, \mu_2, \Phi, \mathbf{r})/d\Phi|_{d\Phi=0}$ are the characteristic functions [2]. We will use the characteristic functions $u_k(\mathbf{r})$ to derive expressions for the electro-chemical capacitance $C_\mu = dQ/d\mu$. In analogy to the electro-chemical capacitance we also derive with the help of the characteristic function v a flux-induced capacitance $C_\Phi = dQ/d\Phi$. For the conductor of Fig. 1(a) which is in an electrically isolating environment an increase in the electro-chemical potential of the reservoir or an increase in flux cannot change the overall charge. For this conductor both the electro-chemical capacitance and the flux-induced capacitance vanish. The conductor of Fig. 1(a) is thus of interest only as a simple limiting case.

For the conductor of Fig. 2(b), the characteristic potentials u_1 and u_2 have the following interesting properties [2]:

(1) For \mathbf{r} deep in reservoir k the local potential must follow the electro-chemical potential of that reservoir and hence $u_k(\mathbf{r}) = 1$.

(2) A change in the electro-chemical potential of reservoir k cannot affect the potential deep inside any other reservoir. Thus for \mathbf{r} deep inside reservoir $l \neq k$ the characteristic function must vanish, $u_k(\mathbf{r}) = 0$.

(3) If we change all electro-chemical potentials simultaneously and by an equal amount $d\mu_k = d\mu$ then we have only changed our (global) energy scale. Hence at every space point \mathbf{r} the potential U must also change by $d\mu$. This implies that the sum of all characteristic functions at every space point is equal to 1,

$$\sum_k u_k(\mathbf{r}) = 1. \quad (2)$$

In contrast to the characteristic potentials u_k which approach 1 in contact k , an increment in flux polarizes the sample but does not affect the state of the system deep inside a reservoir. Hence the characteristic function $v(\mathbf{r})$ vanishes deep inside the electron reservoirs.

To treat the Coulomb interactions it is useful to take the density functional theory as a guide [29, 30]. A discussion of persistent currents based on this approach is given in Ref. [25]. We are interested in the magnitude of the variation of the potential with flux. To this extent we will discuss the closed loop structure in Fig. 2. If the loop is taken to be in an electrically isolating environment the electro-static potential is a function of flux only. (There is no dependence on a chemical potential). The single particle wave functions of the ground state of the ring are thus determined by a Hamiltonian

$$H_{eff} = \frac{1}{2m} \left(\mathbf{p} - \hbar \frac{2\pi}{L} \frac{\Phi}{\Phi_0} \right)^2 + eU_{eff}(\Phi, \mathbf{r}) \quad (3)$$

Here Φ is the flux, $\Phi_0 = h/e$ is the single charge flux quantum, and $L = 2\pi R$ is the circumference of the loop. The effective potential $eU_{eff}(\Phi, \mathbf{r})$ contains in addition to the electro-static potential an exchange potential [29, 30]. The

total kinetic energy of the electrons in the loop is

$$E_{km} = \sum E_n(\Phi, U(\Phi, r)) - \int d^3r n(\Phi, r) e U_{eff}(\Phi, r) \quad (4)$$

where $E_n(\Phi, U_{eff}(\Phi, r))$ are the eigenvalues of eq. (3) and $n(\Phi, r)$ is the electron-density. The sum is over all occupied levels. The persistent current of this loop is the flux-derivative of the free energy. At $kT = 0$ the free energy is equal to the total internal energy $E_{int} = E_{kin} + E_c + E_{ex}$, where E_c is the Coulomb interaction energy and E_{ex} is the exchange and correlation energy [29, 30]. At $kT = 0$ the persistent current of a closed loop is determined by

$$I_{eq}(\Phi) = -c dE_{int}/d\Phi. \quad (5)$$

The persistent current does not in an explicit way depend on the characteristic function $u(\Phi)$ introduced above. This is a consequence of the fact that for potential variations away from the true equilibrium potential the internal energy is stationary [30]. As remarked already in Ref. [25], the total flux derivative in eq. (5) can be replaced by a partial derivative taken at constant potential U . Nevertheless, the potential variation with flux is important: The spectrum of the interacting and the non-interacting system are not the same.

For the open conductors of Fig. 1 it is the grand canonical potential, $\Omega = E_{int} - \sum_k N_k \mu_k$ that counts. The persistent current in these structures is

$$I_{eq}(\Phi) = -c d\Omega/d\Phi. \quad (6)$$

The electro-chemical capacitance coefficients are

$$C_{\mu, k}(\Phi) = -e^2 d^2\Omega/d\mu_k d\mu_l. \quad (7)$$

and the flux-induced capacitances are

$$C_{\Phi, k}(\Phi) = -e d^2\Omega/d\Phi d\mu_k. \quad (8)$$

From this definition of the flux-induced capacitance another alternate interpretation of the second order mixed derivatives of the grand canonical potential becomes apparent. Since $I_{eq}(\Phi) = -c d\Omega/d\Phi$ the flux induced capacitance is related to a "conductance" $G_{\Phi, k} = c C_{\Phi, k} = -ec(dI_{eq}/d\mu_k)$ which is a measure of the sensitivity of the persistent current to a small change in the electro-chemical potential. The flux-induced capacitances are thus connected to the gate voltage dependence of the persistent current.

We make a clear distinction between closed loops with a fixed number of carriers and open loops which are via leads connected to electron reservoirs. Much of the mesoscopic literature attempts to treat closed loops in a grand canonical ensemble and corrects this with a flux dependent chemical potential. Here electro-chemical potentials characterize metallic contacts. For interacting systems it is in addition necessary to distinguish the electrically isolated ring from rings which interact via long range Coulomb forces with other nearby conductors. We use as a reference state a self-consistent equilibrium state. The Coulomb interactions are determined by the actual sample specific charge distribution and not with respect to a flux averaged or ensemble average charge density.

3. Coulomb driven suppression of level hybridization

Consider a one-dimensional loop with a weak disorder potential. A stub, a wire of finite length, is via a barrier

coupled to this loop (see Fig. 2). This structure incorporates some features which are typically encountered in multi-channel rings with finite cross-sections [14, 15]: As a function of flux a single particle energy exhibits regions in which the state is nearly flux-insensitive $dE_n/d\Phi \sim 0$ and behaves as if it were a localized state. These localized regions are interrupted by rapid changes as a function of flux and the state behaves as if it were highly mobile. We can view such a multichannel spectrum as a hybridization of a highly mobile subsystem with a subsystem of localized states. The simple enough example of Fig. 2 allows to investigate the interplay of states which are localized in the stub with highly mobile states in the loop.

We assume that the electro-static potential can be taken uniform inside the stub and can be taken uniform inside the loop. Let U_s denote the electro-static potential [31] (bottom of the conduction band) inside the stub (index s). The electrostatic potential inside the ring (index r) is U_r . If the barrier is not transparent the states in the stub have energy $E_{s, m} + eU_s$, with $m = 1, 2, \dots, M$. The spectrum of the states in the ring is $E_{r, n}(\Phi) + eU_r$, with $n = 1, 2, \dots, N$. A spectrum of such a system is shown in Fig. 3(a). The energies of the flux sensitive (mobile) states of the loop are drawn as solid lines. The energies of the flux-insensitive (localized) states of the stub are shown as broken lines. The N electrons in the loop give a persistent current

$$I_N(\Phi) = -c \sum_{n=1}^N dE_{r, n}(\Phi)/d\Phi \quad (9)$$

shown in Fig. 4.

Now let us make the barrier transparent. Now the stub and the ring form one combined system. Assume that the transparency of the barrier is very small. The spectrum of the combined system will undergo only very small changes where levels of the disconnected system intersect. In a first step we evaluate the spectrum of the combined system taking the potentials U_s and U_r to be known. In a second step we will include the Coulomb interaction to determine these potentials. Consider any two levels E_s and E_r which in the absence of transmission intersect. In the combined

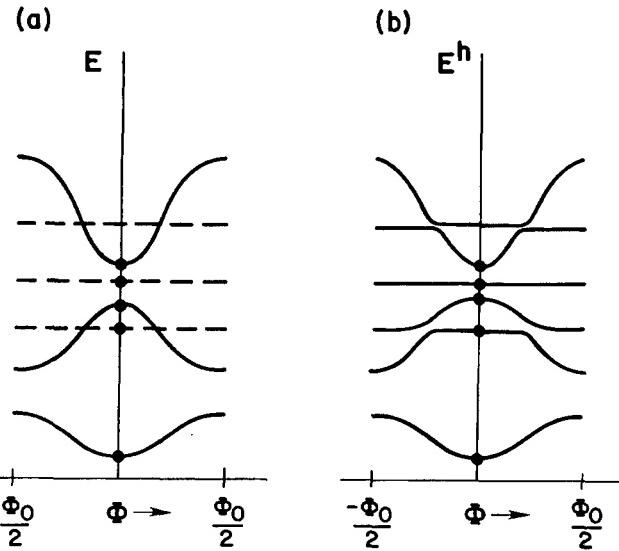


Fig. 3. Energy spectrum as a function of flux for the conductor of Fig. 2. (a) The ring and the stub are completely disconnected. (b) The barrier between the ring and the stub is transparent. Coulomb interactions are not taken into account.

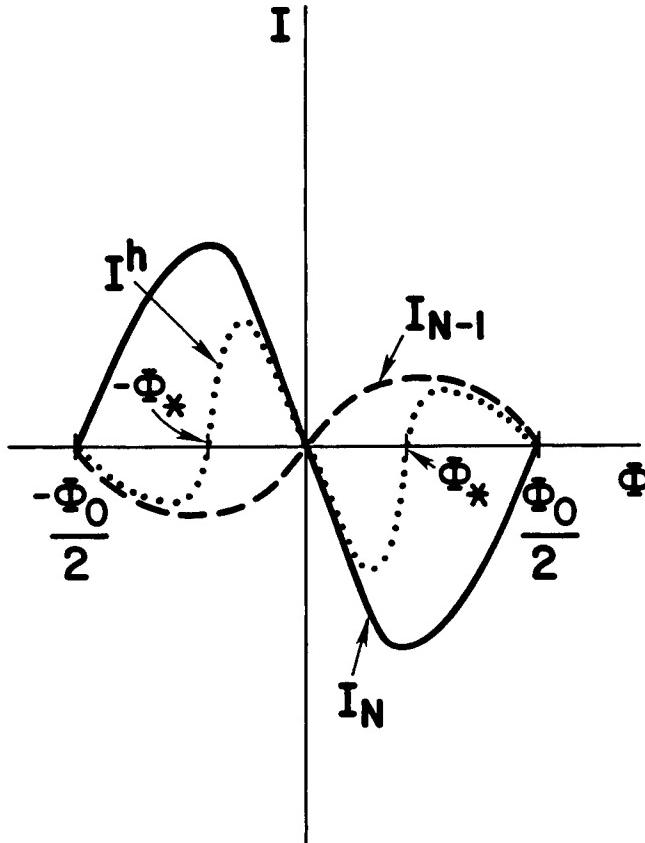


Fig. 4. Persistent current I_N (solid line) and I_{N-1} (dashed line) of the completely disconnected system with N and $N-1$ electrons in the ring. For a transparent barrier and in the absence of interactions the persistent current is I^h (dotted line). Φ_* is the flux at which the topmost energy level of the loop hybridizes with a localized state of the stub. In the presence of interactions the persistent current I_N is recovered even for a transparent barrier.

system the energy levels of these two states have the hybridized energies

$$E_{\pm}^h = (1/2)(E_s + eU_s + E_r + eU_r \pm \Delta) \quad (10)$$

where the energy gap between the two levels is given by

$$\Delta = ((E_s + eU_s - E_r - eU_r)^2 + 4|t|^2)^{1/2}. \quad (11)$$

Here $|t|$ is the energy which couples the states of the ring and the stub. The spectrum E_l^h with $l = 1, \dots, M+N$ of the combined system is shown in Fig. 3(b). It consists of pairs of states with eigenvalues given by eq. (10) and consists of states $E_l^h = E_{s,m} + eU_s$, which do not intersect any level of the loop but fall into a gap. The persistent current of the hybridized system (in the absence of Coulomb interactions) is given by $I^h(\Phi) = -c \sum_l dE_l^h/d\Phi$. In the absence of inter-

actions the flux dependence still originates from the flux dependence of the energies $E_{r,n}$. Since the hybridized states occur, except for the topmost occupied state, in pairs the current is the sum of the persistent current of the uncoupled loop with $N-1$ electrons plus the contribution of the topmost occupied state with energy E_-^h in the hybridized system,

$$I^h(\Phi) = I_{N-1}(\Phi) - c dE_-^h/d\Phi. \quad (12)$$

For a small flux we have $dE_-^h/d\Phi \sim dE_N/d\Phi$ and the persistent current of the combined system is the same as that of the disconnected system. But for a flux far beyond the hybridization point Φ_* we have $dE_-^h/d\Phi \sim 0$. Thus beyond the hybridization point the persistent current is that of an

$N-1$ electron loop. This behavior is shown in Fig. 4. The current of $N-1$ electron states is typically of opposite sign and smaller in magnitude than that of the N electron loop. Thus the current of the hybridized system is typically smaller in magnitude than the persistent current of the decoupled system. Below we will now show that if the Coulomb interactions are taken into account and are of sufficient strength the current of the combined system is in fact restored to a value close to that of the uncoupled system.

The hybridized states with energies E_{\pm}^h are coherent superpositions of a state that was originally localized in the stub and a state that originally was confined to the ring. As a consequence of the superposition the state E_+^h has only a fractional charge Q_{s+} inside the stub and has a fractional charge Q_{r+} inside the ring. Similarly the state with energy E_-^h has a fractional charge Q_{s-} inside the stub and has a fractional charge Q_{r-} inside the ring. These charges can be found by differentiating the energies of these states with respect to the potentials U_s and U_r . With $\alpha = r, s$ we find for the partial charges

$$Q_{\alpha, \pm} = dE_{\pm}^h/dU_{\alpha}. \quad (13)$$

Of course the total charge in each filled state is $Q_{s\pm} + Q_{r\pm} = e$. From eq. (13) we see immediately that if both states \pm are filled then the combined charge of these two states in the stub is $Q_{s+} + Q_{s-} = e$ and in the ring is $Q_{r+} + Q_{r-} = e$. Consequently a net charge motion from the ring into the stub occurs only if the topmost occupied (empty) state in the ring intersects an empty (occupied) state in the stub as shown in Fig. 3(b). Thus the Coulomb interaction is entirely determined by the charge motion in the topmost level E_-^h .

Before we proceed to evaluate the potentials we calculate the variation of the charges of the topmost occupied level E_-^h in response to a small variation of the potentials. There is a Lindhard function $\Pi_{\alpha\beta}$, with $\alpha = r, s$ and $\beta = r, s$ which gives the charge response in the stub or in the ring due to a variation of the potential in the stub or in the ring, $dQ_{\alpha} = -\Pi_{\alpha\beta} dU_{\beta}$. From eq. (13) we find $\Pi_{\alpha\beta} = d^2E_{\pm}^h/dU_{\alpha} dU_{\beta}$. Using eq. (10) we find after a little algebra $\Pi_{ss} = \Pi_{rr} = -\Pi_{rs} = -\Pi_{sr} = -\Pi$ where

$$\Pi = 2e^2 |t|^2 / \Delta^3 \quad (14)$$

In the absence of interactions, i.e. for fixed potentials U_s and U_r , the Lindhard function exhibits a sharp peak of magnitude $1/|t|$ at the point of hybridization and is small as soon as the difference in energies of the two states exceeds $|t|^2$. Next we want to characterize the variation in charge in the stub and in the ring as a function of flux. For fixed potentials an increment $d\Phi$ in the flux causes a change in the charge on the stub given by $(dQ_s/d\Phi)_U = -(d(dE_-^h/dU_s)/d\Phi)_U$. Since at constant potential the flux dependence of the energy stems from $E_{r,N}$ only we find $(dQ_s/d\Phi)_U = -(d^2E_-^h/dU_s dE_{r,N}) (dE_{r,N}/d\Phi)_U$. But $d^2E_-^h/dU_s dE_{r,N} = (1/c) d^2E_-^h/dU_s dU_r$, and hence using eq. (14) we find for the flux-induced charge variation

$$(dQ_s/d\Phi)_U = (1/e)\Pi(dE_{r,N}/d\Phi) \quad (15)$$

Since the total charge in each state is conserved an increment in flux leads to a charge $(dQ_r/d\Phi)_U = -(1/e)\Pi(dE_{r,N}/d\Phi)$ in the ring. We are now ready to determine the self-consistent potentials.

An increment $d\Phi$ in the AB-flux changes the charge in the stub by an amount $(dQ_s/d\Phi)_U d\Phi$. Screening of these charges causes a variation of the potentials by dU_s and dU_r . According to the Lindhard function this causes an additional charge variation in the stub given by $-\Pi dU_s + \Pi dU_r$. The Coulomb interaction is taken into account with the help of a geometrical capacitance C between the stub and the ring. Taking the total charge variation on the stub to be equal to that permitted by the Coulomb interaction gives

$$(dQ_s/d\Phi)_U d\Phi - \Pi dU_s + \Pi dU_r = C(dU_s - dU_r). \quad (16)$$

The charge in the ring obeys an equation which term by term is identical to eq. (16) except for the sign of the charges. Thus eq. (16) determines only the difference in potentials. The solution of eq. (16) is the difference $v = v_s - v_r$ of the characteristic potentials $v_s = e dU_s/d\Phi$, $v_r = e dU_r/d\Phi$,

$$v = e d(U_s - U_r)/d\Phi = \Pi/(C + \Pi)(dE_{r,N}/d\Phi) \quad (17)$$

Integration of eq. (17) gives

$$e(U_s - U_r) = \int^{\Phi} d\Phi' \Pi/(C + \Pi)(dE_{r,N}/d\Phi'). \quad (18)$$

The Lindhard function Π is also a function of the difference of the potentials $e(U_s - U_r)$. Thus eq. (20) is similar to the self-consistent equations encountered in other problems, for instance the BCS gap equation. We note that the Lindhard function modifies the geometrical capacitance and gives rise to a quantum correction. The charge induced in response to an external potential is determined by an effective capacitance [1, 4] $1/C_{eff} = 1/C + 1/\Pi$.

Fortunately there is a simple limiting case in which the solution of the self-consistent equation is obvious. For a very small geometrical capacitance the r.h.s. of eq. (18) is independent of the potentials. The characteristic potential is then determined by the energy derivative of the topmost state of the completely decoupled system, $v(\Phi) = dE_N/d\Phi$. Integration gives

$$e(U_s - U_r) = E_{r,N}(\Phi) - E_{r,N}(0) + eU_0 \quad (19)$$

where U_0 is the potential difference between the conduction band bottoms of the ring and the stub at zero flux. Using eq. (19) in eq. (11) we find that for the topmost occupied level $\Delta = ((E_s - E_{r,N}(0) - eU_0)^2 + 4|t|^2)^{1/2}$ is independent of flux.

Let us now show that in the limit of vanishing capacitance the persistent current of the interacting system is equal to that of the decoupled system. At $kT = 0$ the total energy for the interacting system is the sum of the kinetic energy (see eq. (4)) and the interaction energy [29, 30] E_c ,

$$E_{tot}(\Phi) = \sum_{n=1}^{N+M} E_n^h - Q_s U_s - eQ_r U_r + E_c. \quad (20)$$

Evaluating this energy we find

$$E_{tot}(\Phi) = \sum_{m=1}^M E_{s,m} + \sum_{n=1}^{N-1} E_{r,n}(\Phi) + E_h^h(\Phi) - eU_r(\Phi) - Q(U_s - U_r) + E_c. \quad (21)$$

But for very small capacitance the charge imbalance Q vanishes and the interaction energy is a flux independent constant. Furthermore, the difference of the hybridized energy

of the topmost state and $eU_r(\Phi)$ is $E_h^h(\Phi) - eU_r(\Phi) = E_{r,N}(\Phi) + E_0$ with E_0 a flux independent energy. Up to flux independent terms the total energy is that of the completely decoupled ring. Thus in the limit of small capacitance the persistent current of the interacting system is the same as that of the original, completely decoupled system. This is a consequence of the suppression of level hybridization through Coulomb interaction.

Below we investigate the characteristic potential v for conductors which are open (see Fig. 1), i.e. connected to electron-reservoirs. This permits us to investigate the capacitances and permits us to investigate the effect on the characteristic potential of other nearby metallic bodies.

4. Characteristic potentials for an open normal loop

Consider the open conductors of Fig. 1. We now relate the characteristic potentials for these conductors to electron densities. A variation in the electro-chemical potential $\mu_k = E_{F,k}(\mathbf{r}) + eU(\mathbf{r})$ by $d\mu_k$ can be accomplished in two ways: We can either increase the Fermi energy by $dE_{Fk} = d\mu_k$ or the electric potential by $e dU_k = d\mu_k$. We imagine a two step process: in the first step we increase the chemical potential in reservoir k by $d\mu_k$, keeping the electro-static potential fixed. As a consequence, an additional charge density $(dn(r, k)/dE)_U d\mu_k$ is injected into the conductor. In a second step we switch on the Coulomb interaction. The added charges create an additional induced electrical potential which in turn gives rise to an induced charge distribution $dn_{ind}(r)$. Thus the total change in charge density due to an electro-chemical potential variation in conductor k is $dn_k(r) = (dn(r, k)/dE)_U d\mu_k + dn_{ind,k}(r)$. Similarly, the variation of the electron density due to a change in flux consists of two contributions: We first evaluate the change in electron density keeping the electric potential fixed. In a second step we evaluate the potential due the polarization of the sample caused by the increase in flux and calculate the contribution of this potential due to the variation in charge density, $dn_k(r) = (dn(r, k)/d\Phi)_U d\Phi + dn_{ind,k}(r)$. In the presence of a variation of the electro-chemical potential and a variation in flux we have thus

$$dn_k(r) = \left(\frac{dn_k(r)}{dE} \right)_U d\mu_k + \left(\frac{dn_k(r)}{d\Phi} \right)_U d\Phi + dn_{ind,k}(r). \quad (22)$$

The induced density $dn_{ind,k}(r)$ in conductor k generated by a variation in the electro-static potential $dU(r)$ can be specified by the Lindhard function (or polarization function) $\Pi_k(r, r')$,

$$dn_{ind,k}(r) = - \int d^3r' \Pi_k(r, r') e dU(r') \quad (23)$$

with a potential given by eq. (2). In eq. (23) the integral d^3r' can be taken over all space enclosed by a volume which contains both conductors including a portion of the reservoirs [2]. (This convention applies also to subsequent volume integrals in this work). To obtain an overall charge neutral system the volume has to be chosen so large that no electric field lines penetrate its surface. Equation (23) is just the continuous space analog of the Lindhard function specified by eq. (15) for the closed system. Invariance of the charge distribution under simultaneous changes in all electro-chemical potentials [1, 2] implies that the integral

over the first or the second spatial argument of the Lindhard function is equal to the density of states in conductor k ,

$$(dn_k(r)/dE)_U = \int d^3r' \Pi_k(r, r'). \quad (24)$$

The density response described by the Lindhard function is a consequence of a change in the equilibrium potential. Since an equilibrium density is an even function of magnetic flux, the Lindhard function is also an even function of the magnetic flux, $\Pi_k(\Phi, r', r) = \Pi_k(-\Phi, r', r)$. This has the consequence that the electro-chemical capacitance is an even function of the magnetic flux. We emphasize that these symmetry properties are characteristic for conductors connected to a single reservoir: Reciprocity symmetries apply for multiprobe conductors [2, 4].

In the insulator (index $k = 0$) separating the conductors, a potential variation can polarize the insulator and induce a charge density $dn_{ind, 0}(r) = - \int d^3r' \Pi_0(r, r') e dU(r')$. No external charges reach the insulating region and thus instead of eq. (24) we find that the volume integral of the Lindhard function over either the first or the second argument vanishes, $\int d^3r' \Pi_0(r, r') = \int d^3r' \Pi_0(r', r) = 0$. Note that the Lindhard function which we introduced for the closed loop (see eq. (14)) also has this property.

Next we write down Poisson's equation for the potential U . If we expand U with respect to $d\mu_k$ we find that the characteristic function u_k is determined by [2]

$$\begin{aligned} -\Delta u_k(r) + 4\pi e^2 \int d^3r' \sum_{l=0}^{l=2} \Pi_l(r, r') \mu_k(r') \\ = 4\pi e^2 (dn_k(r)/dE)_U. \end{aligned} \quad (25)$$

Equation (25) contains the sum of all Lindhard functions of all the conductors $l = 1, 2$ and of the insulating region $l = 0$. The density of states of conductor k plays the role of a source term for the characteristic function u_k . Similarly the characteristic function $v(r)$ is a solution of the Poisson equation

$$\begin{aligned} -\Delta v(r) + 4\pi e^2 \int d^3r' \sum_{l=0}^{l=2} \Pi_l(r, r') v(r') \\ = 4\pi e^2 (dn_1(r)/d\Phi)_U. \end{aligned} \quad (26)$$

The flux-sensitivity of the electron density in the conductor with the loop is the source term of the characteristic function v . If the source term is replaced by a test charge $e\delta(r - r_0)$ which is concentrated at one point r_0 the solution to eq. (25) and eq. (26) is Green's function $g(r, r_0)$. With the help of Green's function we find for the characteristic function,

$$u_k(r) = \int d^3r' g(r, r') (dn_k(r')/dE)_U. \quad (27)$$

Similarly the characteristic function $v(r)$ is given by

$$v(r) = \int d^3r' g(r, r') (dn_1(r')/d\Phi)_U. \quad (28)$$

Equation (2) implies for Green's function the property [2]

$$\int d^3r' g(r, r') \sum_k (dn_k(r')/dE)_U = 1. \quad (29)$$

The same relationship follows from the condition that the sum of all induced charge densities plus the test charge is zero. Equation (29) will be used to demonstrate charge and current conservation of the results derived below.

5. Electro-chemical and flux-induced capacitance

The characteristic potentials derived above can now be used to find the electro-chemical capacitance and a flux-induced capacitance. Using eqs. (22), (23), and (27), the total charge in conductor k can be expressed in terms of density of states and Green's function. Differentiating the total charge dQ_k with respect to the voltage $dV_i = d\mu_i/e$ gives an electro-chemical capacitance [2] $C_{ki} = e dQ_k/d\mu_i$ given by

$$\begin{aligned} C_{kl} = e^2 \int d^3r \int d^3r' (dn_k(r)/dE)_U (\delta_{kl} \delta(r - r') \\ - g(r, r') (dn_l(r')/dE)_U). \end{aligned} \quad (30)$$

Equation (30) expresses the capacitances in terms of the density of states of the reference state and Green's function which mediates the Coulomb interactions. Conservation of charge relates the capacitance coefficients of our two-terminal conductor as follows: $C_{11} = C_{22} = -C_{12} = -C_{21}$. To see this one makes use of eq. (29). Below we will use the abbreviation $C_\mu = C_{11}$ where the index μ reminds us that we deal not with an electro-static capacitance but with an electro-chemical capacitance.

Next, in analogy to the electro-chemical capacitance coefficients just discussed we consider the flux induced capacitance $C_{\Phi, k} = dQ_k/d\Phi$. This is the ratio of the piled up charge in conductor k and the increment in flux Φ . We find the flux-induced capacitances

$$\begin{aligned} C_{\Phi, k} = e \int d^3r \int d^3r' (\delta_{1k} \delta(r - r') - (dn_k(r)/dE)_U g(r, r')) \\ \times (dn_1(r')/d\Phi)_U. \end{aligned} \quad (31)$$

These two coefficients are related. Charge conservation implies that the sum of these two coefficients is zero. This can again be demonstrated by using eq. (29). Hence we are left with one coefficient only and denote it by C_Φ , $C_\Phi = C_{\Phi, 1} = -C_{\Phi, 2}$. Whereas the electro-chemical capacitance is an even (and periodic) function of flux the flux-induced capacitance is an odd (and periodic) function of flux.

Consider now for a moment the conductor of Fig. 1(a) which is an electrically neutral environment. Then according to eq. (29) the spatial integral over r of Green's function and the density of states is just equal to 1. Hence in this case both C_μ and C_Φ vanish.

In the presence of time-dependent chemical potentials and in the presence of a time-dependent oscillating flux the current $I_1(t)$ measured at terminal 1 is determined by $I_1(t) = dQ_1/dt$ and the current at terminal 2 (the gate) is determined by $I_2 = dQ_2/dt$. Since charge is conserved, the current is conserved, $I_1(t) + I_2(t) = 0$. The current depends only on the difference of the chemical potentials. Taking the oscillating chemical potential difference to be $d\mu_1 - d\mu_2 = e dV_\omega \exp(-i\omega t)$ and keeping the flux fixed gives rise to a current $dI_\omega = -i\omega C_\mu dV_\omega$. Measurement of this current in a zero-impedance external circuit determines the electro-chemical capacitance. An oscillatory flux with Fourier component $d\Phi_\omega \exp(-i\omega t)$ gives a current $dI_\omega = -i\omega C_\Phi d\Phi_\omega$ in a zero-impedance external circuit.

If the external circuit has a non-vanishing impedance an oscillatory flux will also lead to an oscillating voltage. In this case a total induced current

$$dI_\omega = -i\omega C_\mu dV_\omega - i\omega C_\Phi d\Phi_\omega \quad (32)$$

is generated. The voltage dV_ω depends on the external impedance. For a circuit with impedance $Z(\omega) = -dV_\omega/dI_\omega$ an oscillatory flux generates a voltage

$$dV_\omega = -\frac{i\omega C_\Phi}{(1/Z_\omega) - i\omega C_\mu} d\Phi_\omega. \quad (33)$$

Since we deal with a capacitive effect it is useful to measure the voltage capacitively and take the external impedance to be capacitive $Z(\omega) = 1/(-i\omega C_{ext})$. In this case the measured voltage is independent of frequency and given by

$$dV_\omega = \frac{C_\Phi}{C_\mu + C_{ext}} d\Phi_\omega. \quad (34)$$

Our discussion of the conductor in Fig. 2 has shown that the charge displaced by the flux is at best one electronic charge and is less if interactions are taken into account. If this is indicative also of the open conductor in Fig. 1 then the magnitude of the flux-induced capacitance is of the order of e/Φ_0 or $C_\Phi = e^2/hc$. This is equivalent to a "conductance" $G_\Phi = cC_\Phi = e^2/h$ equal to the fundamental conductance unit. Screening will reduce these values but both the measurement of the electro-chemical capacitance and the flux-induced capacitance should be feasible.

In this work we have discussed the response of the internal electro-static potential of mesoscopic conductors to small changes in an external parameter (flux, electro-chemical potential) with the help of characteristic functions. The characteristic potential v is responsible for the enhancement of the persistent current above a value achieved in the non-interacting system. This potential also determines a flux-induced capacitance. The characteristic function u determines the electro-chemical capacitance. Measurement of these capacitances can, therefore, provide an experimental means to determine the sensitivity of the electric potential.

Note added in proof

To obtain an enhancement of the persistent current in situations different from that shown in Fig. 3 it is in general necessary that the stub acts like an *acceptor* in the hybridized system. The stub acts like an acceptor (rather than a donor) if its density of states is much larger than that of the loop. This corresponds to the situation in diffusive multi-channel rings where we have many localized states and only a few highly mobile states.

Acknowledgements

I would like to acknowledge the support of the SERC and the hospitality of the Cavendish Laboratory in Cambridge, U.K., where part of this work was done. I also would like to acknowledge instructive discussions with N. D. Lang which helped to correct an earlier version of this manuscript.

References

1. Büttiker, M., Thomas, H. and Prêtre, A., Phys. Lett. **A180**, 364 (1993); Büttiker, M., Prêtre, A. and Thomas, H., Z. Phys. – Condens. Matter **B94**, 133 (1994).
2. Büttiker, M., J. Phys.: Condens. Matter **5**, 9631 (1993).
3. Fomin, N. V., Sov. Phys. – Solid State **9**, 474 (1967); Stern, F., "Low Temperature Capacitance of Inverted Silicon MOS Devices in High Magnetic Fields", IBM Research Report, RC 3758, March 2 (1972).
4. Chen, W., Smith, T., Büttiker, M. and Shayegan, M., Phys. Rev. Lett. **73**, 146 (1994).
5. Büttiker, M., Imry, Y. and Landauer, R., Phys. Lett. **A96**, 365 (1983).
6. Pauling, L., J. of Chem. Phys. **4**, 673 (1936).
7. London, F. J., Phys. Radium **8**, 397 (1937).
8. Hund, F., Ann. Phys. (Leipzig), **32**, 102 (1938).
9. Beyers, N. and Yang, C. N., Phys. Rev. Lett. **46**, 7 (1961).
10. Bloch, F., Phys. Rev. **137**, 787 (1965).
11. Landauer, R. and Büttiker, M., Phys. Rev. Lett. **54**, 2049 (1985).
12. Büttiker, M., Phys. Rev. **B32**, 1846 (1985).
13. Büttiker, M. and Klapwijk, T. M., Phys. Rev. **B33**, 5114 (1986).
14. Cheung, H. F., Gefen, Y. and Riedel, E. K., IBM J. Res. Develop. **32**, 379 (1988).
15. Bouchiat, H. and Montambaux, G., J. Phys. (Paris) **50**, 2695 (1989).
16. Levy, L. P., Dolan, G., Dunsmuir, J. and Bouchiat, H., Phys. Rev. Lett. **64**, 2074 (1990).
17. Chandrasekhar, V., Webb, R. A., Brady, M. J., Ketchen, M. B., Gallagher, W. J. and Kleinsasser, A., Phys. Rev. Lett. **67**, 3578 (1991).
18. Mailly, D., Chapelier, C. and Benoit, A., Phys. Rev. Lett. **70**, 2020 (1993).
19. Altshuler, B. L., Gefen, Y. and Imry, Y., Phys. Rev. Lett. **66**, 88 (1991).
20. von Oppen, F. and Riedel, E. K., Phys. Rev. Lett. **66**, 84 (1991).
21. Akkermans, E., Auerbach, A., Avron, J. E. and Shapiro, B., Phys. Rev. Lett. **66**, 76 (1991).
22. Schmid, A., Phys. Rev. Lett. **66**, 80 (1991); Ambegaokar, V. and Eckern, U., Phys. Rev. Lett. **65**, 381 (1990).
23. Loss, D. and Martin, Th., Phys. Rev. **B47**, 4619 (1993).
24. Kopietz, P., Phys. Rev. Lett. **70**, 3123 (1993); G. Vignale, Phys. Rev. Lett. **72**, 433 (1994).
25. Argaman, N. and Imry, Y., Physica Scripta **T49**, 333 (1993).
26. Müller-Groeling, A. and Weidenmüller, H. A., Phys. Rev. **B49**, 4752 (1993).
27. Mello, P., Phys. Rev. **B47**, 16358 (1993).
28. Singha Deo, P. and Jayannavar, A. M., Mod. Phys. Lett. **B7**, 1045 (1993); Jayannavar, A. M. and Singha Deo, P., Phys. Rev. **B49**, 13685 (1994).
29. Kohn, W. and Sham, L. J., Phys. Rev. **140**, A1133 (1965); Hohenberg, P. and Kohn, W., Phys. Rev. **136**, B864 (1964).
30. Lang, N. D., in: "Theory of the Inhomogeneous Electron Gas" (Edited by S. Lundquist and N. H. March) (Plenum Press, New York 1983), p. 309.
31. Throughout most of this work we make no distinction between U_{eff} and U and thus neglect exchange and correlation effects.

Raman Scattering from a Circular Quantum Dot

A. G. Mal'shukov

Institute of Spectroscopy, Russian Academy of Sciences, 142092 Troitsk, Moscow Region, Russia

and

Arne Brataas and K. A. Chao

Division of Physics, Faculty of Physics and Mathematics, Norwegian Institute of Technology, University of Trondheim, 7034 Trondheim, Norway

Received April 28, 1994; accepted June 14, 1994

Abstract

Electronic Raman scattering from a doped circular quantum dot is studied within the random phase approximation. The Raman spectrum of the spin density fluctuation channel consists of only sharp peaks due to single-particle excitations. However, for the charge density fluctuation channel, the broadened plasmon peak in a small quantum dot may overlap with the sharp single-particle peaks, resulting in fine structures in the Raman spectrum.

1. Introduction

Raman scattering is a useful tool for studying the electronic spectra of semiconductors [1]. The Raman spectrum of a system depends on the characteristic features of both the single-particle and the plasmon excitations. In a quantum dot of the size about 1000 Å, the single-particle excitation energies are discrete, resulting in sharp peaks in Raman spectra. On the other hand, the plasmon frequency is also influenced by the size of system because it is a collective excitation of electron-hole pairs. In particular, the broadening of the plasmon peak is of the order v_f/a [2, 3], where v_f is the Fermi velocity and a the dimension of the dot. Therefore, in a small quantum dot it is possible that the broadened plasmon mode may overlap with the sharp single-particle peaks. When this occurs, fine structures appear in the plasmon excitation. In this paper the Raman spectrum of a perfect quantum dot will be calculated with the random phase approximation (RPA).

2. Theory

Consider the Raman scattering from a semiconductor quantum dot. The effective interaction Hamiltonian between electrons and the electromagnetic wave can be written as [4]

$$H_{int} = \gamma_{cd} |\mathbf{e}_i \cdot \mathbf{e}_s| [n_\uparrow(\mathbf{q}) + n_\downarrow(\mathbf{q})] + i\gamma_{sd} |\mathbf{e}_i \times \mathbf{e}_s| [n_\uparrow(\mathbf{q}) - n_\downarrow(\mathbf{q})] \hat{a}_s^\dagger \hat{a}_i + \text{h.c.}, \quad (1)$$

where \mathbf{q} is the difference of wave vectors between the incoming (represented by the operator \hat{a}_i) and the scattered wave (represented by the operator \hat{a}_s^\dagger), and [5]

$$\gamma_{cd} = r_0 \left[1 + \frac{2p^2}{3m} \sum_\alpha \frac{\Delta_\alpha}{\Delta_\alpha^2 - \omega_i^2} \right], \quad (2)$$

$$\gamma_{sd} = \frac{p^2 \omega_i}{3m} \sum_\alpha \frac{s_\alpha}{\Delta_\alpha^2 - \omega_i^2}. \quad (3)$$

Here $r_0 = e^2/(mc^2)$ is the classical electron radius. $\alpha = 1, 2$ and 3 specify, respectively, the heavy, the light and the split-off valence band. Δ_α is the energy gap between the conduction and the α th valence band (with $\Delta_1 = \Delta_2$), and $s_1 = s_2 = 1$ and $s_3 = -2$. $p = \langle \mathbf{x} | p_x | s \rangle$ is the momentum matrix element between the s -like conduction band and the valence band which transforms like x . We should notice that (1) with the coupling constants (2) and (3) is valid only when the incident and the scattered photon frequencies are not very close to Δ_1 or to Δ_3 .

The two terms in H_{int} represent the charge density fluctuation (CDF) and the spin density fluctuation (SDF). Under the scattering geometries $\mathbf{e}_i \parallel \mathbf{e}_s$ and $\mathbf{e}_i \perp \mathbf{e}_s$, the CDF and the SDF give, respectively, the Raman scattering cross sections as

$$\frac{d\sigma_{cd}}{d\Omega d\omega} = r_0^2 \gamma_{cd}^2 (\mathbf{e}_i \cdot \mathbf{e}_s)^2 \frac{1}{\pi} S_{cd}(\omega, q), \quad (4)$$

$$\frac{d\sigma_{sd}}{d\Omega d\omega} = r_0^2 \gamma_{sd}^2 (\mathbf{e}_i \times \mathbf{e}_s)^2 \frac{1}{\pi} S_{sd}(\omega, q), \quad (5)$$

where S_{cd} and S_{sd} are the corresponding spectral functions.

To obtain S_{cd} and S_{sd} , we need to calculate the polarisation operator. Within the RPA, it can be expressed as

$$\Pi(\mathbf{r}, \mathbf{r}'; \omega) = \frac{e^2}{\kappa} \sum \frac{\psi_\mu(\mathbf{r}) \psi_\nu^*(\mathbf{r}) \psi_\mu(\mathbf{r}') \psi_\nu^*(\mathbf{r}')}{\varepsilon_\mu - \varepsilon_\nu - \omega - i\delta}, \quad (6)$$

where κ is the dielectric constant of the surrounding medium and $\psi_\mu(\mathbf{r})$ is the single-particle wave function with eigenenergy ε_μ . We consider the case that the dipole approximation, $qa \ll 1$ is valid. Then, for the SDF channel, RPA gives

$$S_{sd}(\omega, q) = q^2 \text{Im} \left[\sum_{\nu, \mu} \frac{|x_{\nu, \mu}|^2 (f(\varepsilon_\nu) - f(\varepsilon_\mu))}{\varepsilon_\nu - \varepsilon_\mu - \omega - i\delta} \right], \quad (7)$$

where $x_{\nu, \mu} = \langle \psi_\nu | \mathbf{x} | \psi_\mu \rangle$. On the other hand, there is no simple expression for $S_{cd}(\omega, q)$. In the dipole approximation $S_{cd}(\omega, q)$ is the imaginary part of the dipole polarisability of the quantum dot in a homogeneous electric field E_{ext} along the direction of \mathbf{q} (parallel to x -axis), and can be written in the general form

$$(3) \quad S_{cd}(\omega, q) = q^2 \text{Im} \left[\frac{P_x}{E_{ext}} \right]. \quad (8)$$

The problem now is to calculate the dipole moment P_x of the dot in a homogeneous field.

To go further, we must specify the quantum dot in order to solve for the eigensolutions. Here we consider a perfect circular dot of radius a , surrounded by an infinite potential barrier. The position in the dot is labelled by $\rho = (\rho, \phi)$. The normalised single-particle eigenfunctions are

$$\psi_\mu^m(\rho, \phi) = c_\mu^{|m|} J_{|m|}(k_\mu^{|m|} \rho) \exp(-im\phi), \quad (9)$$

where $J_{|m|}$ is the Bessel function of the first kind. $m = 0, \pm 1, \pm 2, \dots$ and $\mu = 0, 1, 2, \dots$ are angular and radial quantum numbers, respectively. In terms of these functions, the electrical dipole moment is given by

$$X_{\mu, v}^m = \int_0^a d\rho \psi_\mu^{m+1}(\rho) \psi_v^m(\rho) \rho^2, \quad (10)$$

and from (7) we readily obtain the low temperature result as

$$S_{sd}(\omega, q) = \frac{\pi q^2}{2} \sum_{\epsilon_\mu^{m+1} < 0, \epsilon_v^m > 0} (X_{\mu, v}^m)^2 \delta(\omega_{\mu, v} - \omega), \quad (11)$$

where $\omega_{\mu, v}^m = \epsilon_v^m - \epsilon_\mu^{m+1}$ is the single-particle excitation energy. The Raman spectrum for the $e_i \parallel e_s$ scattering geometry then consists of sharp peaks, since in RPA treatment for the SDF channel the electron correlation is neglected.

For the CDF channel in the $e_i \perp e_s$ scattering geometry the problem is more complicated, because of the induced charge $n_{ind}(\mathbf{r})$ and the induced potential $\phi_{ind}(\mathbf{r})$ which are functions of $\mathbf{r} = (\rho, z)$. While the induced potential is given by the induced charge

$$\phi_{ind}(\mathbf{r}) = \int d\mathbf{r}' \frac{1}{|\mathbf{r} - \mathbf{r}'|} n_{ind}(\mathbf{r}'), \quad (12)$$

the induced charge itself can be expressed in terms of the total potential $\phi(\mathbf{r}) = \phi_{ind}(\mathbf{r}) + \phi_{ext}(\mathbf{r})$ as

$$n_{ind}(\mathbf{r}) = \int d\mathbf{r}' \Pi(\mathbf{r}, \mathbf{r}'; \omega) \phi(\mathbf{r}'). \quad (13)$$

If we introduce the surface charge density

$$n_s(\rho) = \int_{-\infty}^{\infty} dz n_{ind}(\mathbf{r}) \quad (14)$$

and surface polarisation operator

$$\Pi_s(\rho, \rho') = \int dz dz' \Pi(\mathbf{r}, \mathbf{r}'), \quad (15)$$

then from (13) and (12) we obtain

$$n_s(\rho) = \int d^2 \mathbf{p}' d^2 \mathbf{p}' \frac{\Pi_s(\rho, \rho')}{|\rho' - \rho'|} n_s(\rho) + \int d^2 \mathbf{p}' \Pi_s(\rho, \rho') \phi_{ext}(\mathbf{p}'). \quad (16)$$

Since the quantum dot has circular symmetry, we can perform the expansions

$$n_s(\rho) = n_s(\rho) \exp(i\phi), \quad (17)$$

$$\Pi_s(\rho, \rho') = \frac{1}{(2\pi)^2} \sum_m \Pi_m(\rho, \rho') \exp(im(\phi - \phi')) \quad (18)$$

Substituting (18) and (17) into (16), and integrating over the angular variable, we arrive at

$$n_s(\rho) = \int d\rho' \int d\rho'' \Pi_1(\rho, \rho'') \Phi(\rho'', \rho) n_s(\rho') \rho' \rho'' + \frac{1}{2\pi} \int d\rho' \Pi_1(\rho, \rho') \rho'^2, \quad (19)$$

where

$$\Phi(\rho, \rho') = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\phi \exp(i\phi)}{\sqrt{\rho^2 + \rho'^2 - 2\rho'\rho \cos\phi}}. \quad (20)$$

To obtain these results, we have considered the dipole solution of (16), for which one can use the convenient form $\phi_{ext}(\mathbf{p}) = \rho \exp(i\phi)$. At low temperature Π_1 reduces to

$$\Pi_1(\rho, \rho') = -\frac{2e^2}{\kappa} \sum_{\epsilon_\mu^{m+1} < 0, \epsilon_v^m > 0} \frac{2\Psi_{\mu, v}^m(\rho) \Psi_{\mu, v}^{m'}(\rho') \omega_{\mu, v}^m}{(\omega_{\mu, v}^m)^2 - (\omega + i\delta)^2}, \quad (21)$$

where $\Psi_{\mu, v}^m(\rho) = \psi_\mu^{m+1}(\rho) \psi_v^m(\rho)$.

The coupled eqs (19) and (21) can be solved via the matrix representation. Let us first define

$$Y_{\mu, v}^m = \left\{ \int d\rho' \int d\rho \Psi_{\mu, v}^m(\rho') \Phi(\rho, \rho') n_s(\rho) \rho' \rho + \frac{1}{2\pi} \int d\rho \rho^2 \Psi_{\mu, v}^m(\rho) \right\} \frac{\sqrt{2\omega_{\mu, v}^m}}{(\omega_{\mu, v}^m)^2 - (\omega + i\delta)^2} \quad (22)$$

and

$$c_{\mu, v, \mu', v'}^{m, m'} = -\frac{e^2}{\kappa} \int_0^a d\rho d\rho' \Psi_{\mu, v}^m(\rho) \Phi(\rho, \rho') \Psi_{\mu', v'}^{m'}(\rho') \rho \rho'. \quad (23)$$

To simplify the notation, we introduce a single parameter α (or β or γ) to represent the three indices (m, μ, v) in all variables having m for superscript and (μ, v) for subscripts. Let us further define a matrix A with elements

$$\sum_\beta A_{\alpha, \beta} z_{\beta, \gamma} = (\Omega_\gamma)^2 z_{\alpha, \gamma}. \quad (24)$$

Then, it can be proved that to solve the two coupled eqs (19) and (21) is equivalent to solving the eigenvalue problem of the matrix A

$$\sum_\beta A_{\alpha, \beta} z_{\beta, \gamma} = (\Omega_\gamma)^2 z_{\alpha, \gamma}. \quad (25)$$

After obtaining the eigensolutions, the spectral function S_{cd} can be expressed as

$$S_{cd} = \frac{q^2 \pi}{2} \sum_{\alpha, \beta, \gamma} x_\alpha \sqrt{\omega_\alpha} z_{\alpha, \beta} \frac{1}{\Omega_\beta} \delta(\Omega_\beta - \omega) z_{\beta, \gamma}^{-1} x_\gamma \sqrt{\omega_\gamma}. \quad (26)$$

The positions of the peaks in the Raman spectra are determined by the Stokes shifts which are just the elementary excitation energies Ω_α . The spectral weight of each peak is given by

$$W_\alpha = \frac{\pi}{\Omega_\alpha} \left(\sum_\beta \sqrt{\omega_\beta} x_\beta z_{\beta, \alpha} \right)^2. \quad (27)$$

3. Numerical results

Based on the analytical expressions derived above, we have calculated the Raman spectra of 2D circular GaAs quantum dots of radius a and 2D electron density n . The material parameters for GaAs are $m^* = 0.067m_e$ and $\kappa = 12.4$. We

have studied dots with radius in the range from $a = 500 \text{ \AA}$ to $a = 1000 \text{ \AA}$, and the electron density from $n = 10^{11} \text{ cm}^{-2}$ to $n = 10^{12} \text{ cm}^{-2}$. When a is reduced the separation between adjacent single-particle levels increases, and the Stokes shift of the Raman scattering due to the SDF channel is enhanced. If the enhancement is sufficiently large, the electron-hole excitations will overlap with the plasmon mode. When this happens, the Raman spectrum due to the CDF will exhibit single-particle oscillation superimposed on the broad plasmon peak. We will illustrate this interesting behaviour with our numerical results.

Figure 1 shows the complete Raman spectrum for a dot of larger radius $a = 1000 \text{ \AA}$ with $n = 9.17 \cdot 10^{11} \text{ cm}^{-2}$. The overlap between the SDF part and the CDF part is extremely small, and the CDF part is due to a conventional plasma excitation. The separation between the SDF part and the CDF part is caused by the long-range Coulomb interaction.

When the size of the dot is reduced to $a = 500 \text{ \AA}$, the Raman spectra for $n = 5.86 \cdot 10^{11} \text{ cm}^{-2}$ are shown in Fig. 2 for the $e_i \parallel e_s$ scattering geometry (the SDF channel), and in Fig. 3 for the $e_i \perp e_s$ scattering geometry (the CDF channel). The CDF spectrum overlaps very much with the SDF spectrum, and so in the CDF spectrum we see quantum oscillation over the plasma excitation at energy $\Omega_6 = 10.86 \text{ meV}$ (Mode 6 in Fig. 3).

To demonstrate the characteristic features of the collective excitation modes in the CDF Raman spectrum, in Fig. 4 we analyze mode 2 (solid line) and mode 6 (dashed line) of Fig. 3. Each collective excitation mode can be expressed as a linear combination of the electron-hole excitation modes, which are enumerated according to the increasing electron-hole excitation energy. The square of the coefficient of linear

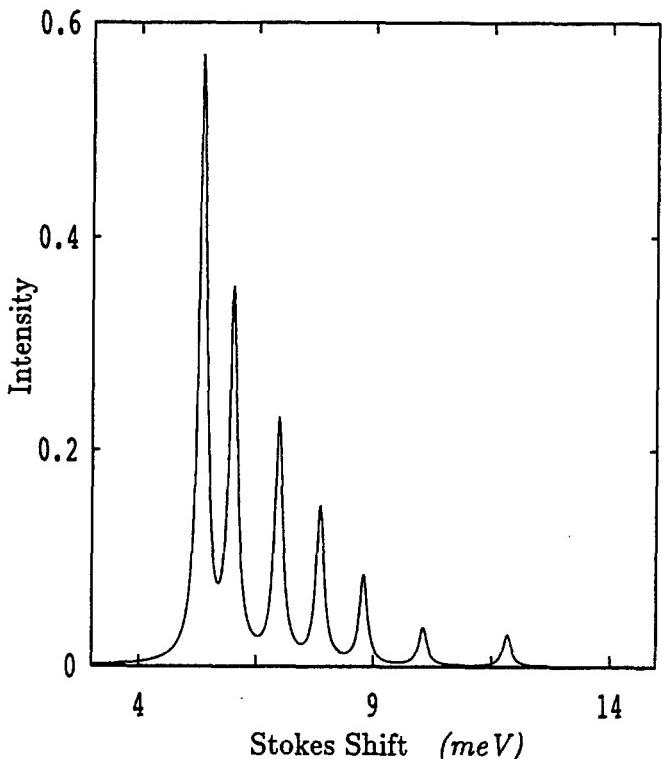


Fig. 2. The SDF-channel Raman spectrum of a 2D circular GaAs quantum dot of radius 500 \AA , and electron density $5.86 \cdot 10^{11} \text{ cm}^{-2}$. The spectrum is normalised to 1.

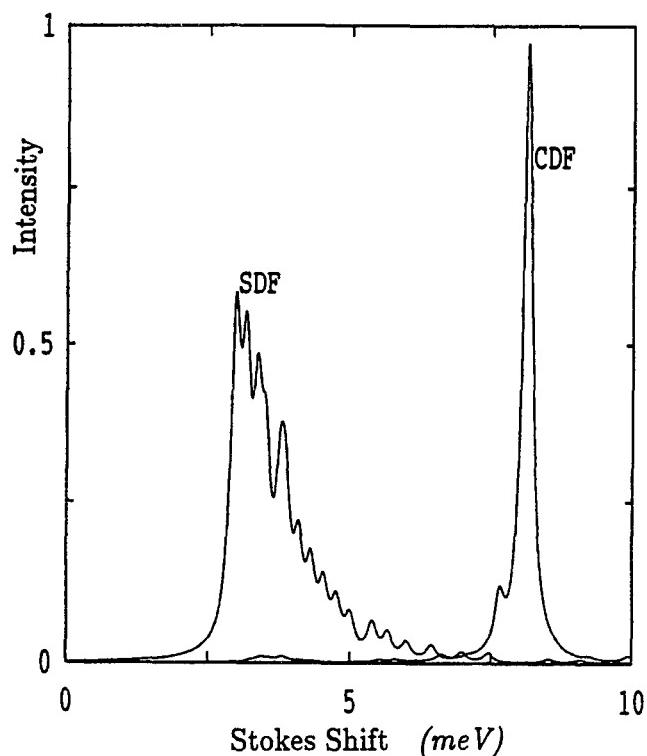


Fig. 1. The Raman spectra of a 2D circular GaAs quantum dot of radius 1000 \AA , and electron density $9.17 \cdot 10^{11} \text{ cm}^{-2}$. The spectrum due to the spin density fluctuation (SDF) channel is separated from that due to the charge density fluctuation (CDF) channel. Each spectrum is normalised to 1.

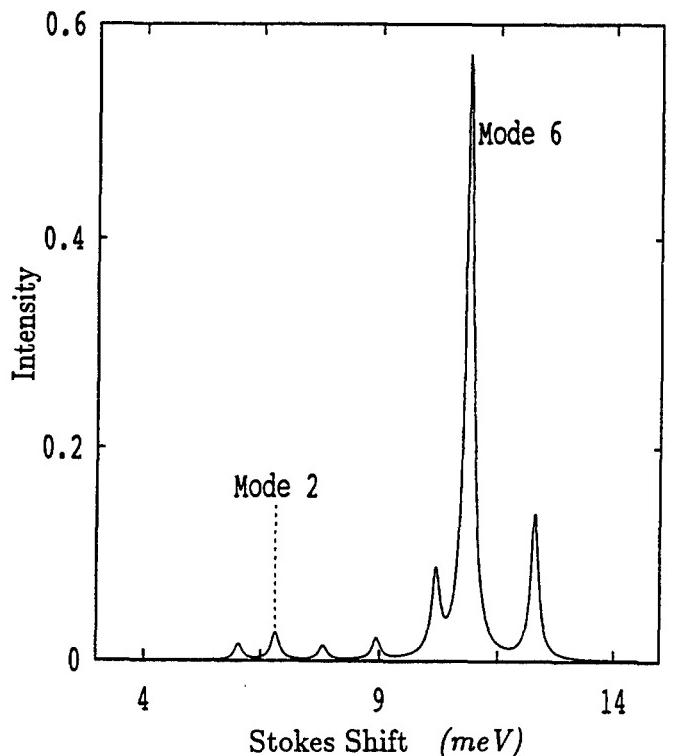


Fig. 3. The CDF-channel Raman spectrum of a 2D circular GaAs quantum dot of radius 500 \AA , and electron density $5.86 \cdot 10^{11} \text{ cm}^{-2}$. The spectrum is normalised to 1.

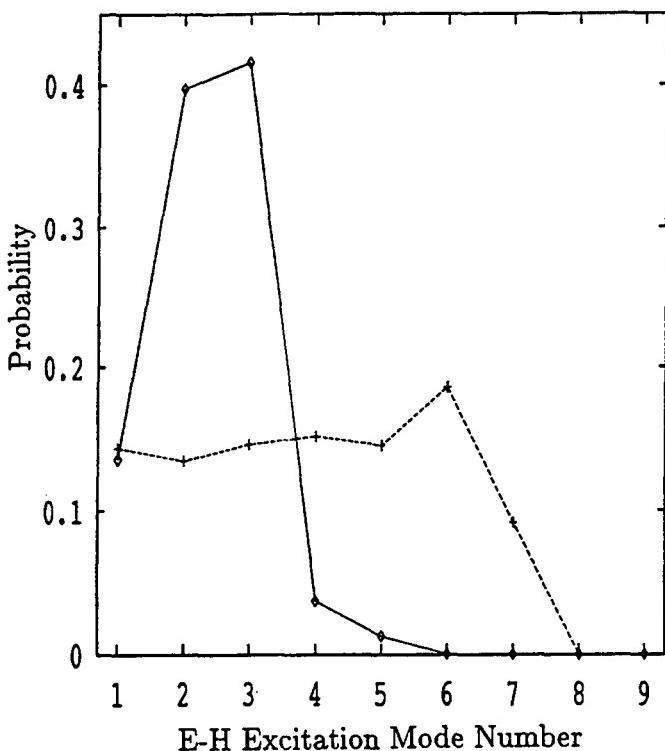


Fig. 4. The eigenvectors of the mode 2 (solid line) and the mode 6 (dashed line) in the CDF-channel Raman spectrum shown in Fig. 3.

4. Final remark

We would like to mention two points relevant to our present work. First, it has been found [6] that the correlation effect is important in an a small quantum dot. Second, the quantum oscillation appeared in the plasma mode remains even if the shape of the quantum dot is irregularly rough, as can be shown with the approach of Dyson ensembles [7].

Acknowledgement

We acknowledge the support from Nordisk Forskerutdanningsakademi and the Norwegian Supercomputing Committee (TRU).

References

1. Pinczuk A. and Burstein, E., in: "Light Scattering in Solids I, Topics Appl. Phys. Vol. 8" (Edited by M. Cardona) (Springer-Verlag, Berlin 1975).
2. Lozovik, Yu. E. and Nishanov, V. N., Sov. Phys. Solid State **20**, 12 (1978).
3. Tran Thoai, D. B. and Ekardt, W., Solid State Commun. **41**, 9 (1982).
4. Hamilton, D. and McWhorter, A., in: "Light Scattering in Solids" (Edited by G. B. Wright) (Springer, New York 1969).
5. Klein, M., in: "Light Scattering in Solids I, Topics Appl. Phys. Vol. 8" (Edited by M. Cardona) (Springer, Berlin 1983).
6. Pfannkuche, D., Gudmundsson, V. and Maksym, P. A., Phys. Rev. **B47**, 2244 (1993).
7. Mal'shukov, A. G., Brataas, A. and Chao, K. A. (unpublished).

Local Field Calculation for a Spherical Semiconductor Quantum Dot with Parabolic Confinement

O. Keller and T. Garm

Institute of Physics, University of Aalborg, Pontoppidanstraede 103, DK-9220 Aalborg Øst, Denmark

Received May 2, 1994; accepted June 15, 1994

Abstract

A self-consistent calculation of the local electromagnetic field inside a spherical quantum dot is presented. The confinement potential is taken to be parabolic and isotropic, and the interaction between the field and the quantum dot electrons is described by a nonlocal paramagnetic response function. A particular feature of the treatment is that the calculation is carried out without making use of the electric dipole approximation. As a consequence the magnitude of the electromagnetic field has a finite value at all positions in space. A resonance condition is set up for the local field, and an expression for the resonance frequency is obtained. It is found that the resonance frequency is blue-shifted with respect to the electronic transition frequency. The theory is applied to a GaAs quantum dot embedded in a $\text{Ga}_{1-x}\text{Al}_x\text{As}$ medium. For this system the local field factor is calculated, and the frequency shift is evaluated.

1. Introduction

The interest in the electromagnetic properties of quantum dots or “solid-state atoms” has increased dramatically with the recent development of nanometer lithographic techniques. One way of realizing dot structures is by means of local interdiffusion in a semiconducting medium, hereby producing extremely localized electron and hole states. Quantum dots have been investigated extensively using e.g. photoluminescence [1] or magnetospectroscopy [2], and experimental as well as theoretical reports on their nonlinear properties have been presented [3–4]. In addition, experimental [5] and theoretical [6] studies of the radiative decay rate of excitons in quantum dots have been published recently.

The spatial extent of the electron and hole wavefunctions are often significantly smaller than the wavelength of the electromagnetic field, and so the (electric) dipole approximation gives a satisfactory description of the far-field properties. However, it is well known that conceptual difficulties, such as a singularity in the near-field, are connected with the calculation of the electric field at the site of an atom embedded in a homogenous medium, when the electric dipole approximation is adopted. In addition a rigorous treatment of the local field will provide a resonance condition, which will be absent within the framework of the dipole approximation. This in turn indicates that the observed resonances in e.g. absorption spectra are shifted from the electronic transition frequencies.

In this paper the local field is calculated without making use of the electric dipole approximation. The calculation is based on a self-consistent scattering formalism, in which the field is obtained from the solution of an integral equation. The nonlocal nature of the interaction between the electrons and the electromagnetic field is taken into account, so that

the dependence on the wavefunctions is made clear. An isotropic parabolic potential is used as a model for the confinement, so that the wavefunctions are those of the well known harmonic oscillator. We restrict ourselves to the case of a single electron quantum dot, and only conduction intraband transitions are considered. The only states which are taken into account are (in hydrogen-like notation): $1s$, $2p_x$, $2p_y$, and $2p_z$, and consequently the quantum dot is regarded as a 4-level system in which the three excited states are degenerate in energy.

2. General formalism

The local electric field, in the frequency (ω) domain, is calculated using the formalism developed in [7] and reviewed in [8]. This formalism is based on an integral equation which relates the field $\mathbf{E}(\mathbf{r})$ at space point \mathbf{r} to the Fourier transformed current density $\mathbf{j}(\mathbf{q})$ as follows:

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) - i\mu_0\omega \frac{1}{(2\pi)^3} \int \mathbf{G}(\mathbf{q}) \cdot \mathbf{j}(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}} d^3q, \quad (1)$$

where $\mathbf{E}^0(\mathbf{r})$ is the background field and $\mathbf{G}(\mathbf{q})$ is the Green's function in Fourier space for the geometry in question. In the following the nonlinear and diamagnetic contributions to the current density are neglected. This implies that the current density in physical space $\mathbf{j}(\mathbf{r})$ is related to the electric field through the nonlocal relation [9]

$$\mathbf{j}(\mathbf{r}) = \int \boldsymbol{\sigma}(\mathbf{r}, \mathbf{r}') \cdot \mathbf{E}(\mathbf{r}') d^3r', \quad (2)$$

where the conductivity tensor $\boldsymbol{\sigma}(\mathbf{r}, \mathbf{r}')$ for real wavefunctions is given by

$$\boldsymbol{\sigma}(\mathbf{r}, \mathbf{r}') = -\frac{ia}{\mu_0\omega} \{ j_{1x}(\mathbf{r})j_{1x}(\mathbf{r}') + j_{1y}(\mathbf{r})j_{1y}(\mathbf{r}') + j_{1z}(\mathbf{r})j_{1z}(\mathbf{r}') \}. \quad (3)$$

Here

$$j_{1a} = -\frac{e\hbar}{2im} \{ \phi_{2pa} \nabla \phi_{1s} - \phi_{1s} \nabla \phi_{2pa} \}, \quad (4)$$

denotes the transition current density for the transition $1s$ to $2p_a$ and the quantity a is given by

$$a = \mu_0 \frac{\rho_{11}^F - \rho_{22}^F}{\hbar(\omega + iv) + E_1 - E_2} + \mu_0 \frac{\rho_{22}^F - \rho_{11}^F}{\hbar(\omega + iv) + E_2 - E_1}, \quad (5)$$

where ρ_{mm}^F is the matrix element of the thermal equilibrium density matrix operator, v is the relaxation frequency, and E_m is the energy eigenvalue for the state m ($m = 1, 2$). After some manipulations it is realized that the solution to the

integral equation may be written [7, 8]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^0(\mathbf{r}) - \frac{a}{1+aN} \{ \gamma_{1x}^0 \mathbf{F}_{1x}(\mathbf{r}) + \gamma_{1y}^0 \mathbf{F}_{1y}(\mathbf{r}) + \gamma_{1z}^0 \mathbf{F}_{1z}(\mathbf{r}) \}, \quad (6)$$

where the following notation is used

$$\mathbf{F}_{1x}(\mathbf{r}) = \frac{1}{(2\pi)^3} \int \mathbf{G}(\mathbf{q}) \cdot \mathbf{j}_{1x}(\mathbf{q}) e^{i\mathbf{q} \cdot \mathbf{r}} d^3q, \quad (7)$$

$$N = \int \mathbf{F}_{1x}(\mathbf{r}) \cdot \mathbf{j}_{1x}(\mathbf{r}) d^3r, \quad (8)$$

and

$$\gamma_{1x}^0 = \int \mathbf{E}^0(\mathbf{r}) \cdot \mathbf{j}_{1x}(\mathbf{r}) d^3r. \quad (9)$$

Here $\mathbf{j}_{1x}(\mathbf{q})$ is the Fourier transformed transition current density. The symmetry properties of the model imply that the functions $\mathbf{F}_{1y}(\mathbf{r})$ and $\mathbf{F}_{1z}(\mathbf{r})$ immediately can be obtained once the explicit expression for $\mathbf{F}_{1x}(\mathbf{r})$ is known. Thus, in the following we shall pay particular attention to this function.

3. Isotropic harmonic oscillator

The two states $1s$ and $2p_x$ are taken as the ground state and first excited state of a 3-dimensional, isotropic harmonic oscillator. The wavefunctions for these states are

$$\phi_{1s}(\mathbf{r}) = \frac{\beta^{3/2}}{\pi^{3/4}} e^{-\beta^2 r^2/2}, \quad (10)$$

and

$$\phi_{2px}(\mathbf{r}) = \sqrt{2} \frac{\beta^{5/2}}{\pi^{3/4}} x e^{-\beta^2 r^2/2}, \quad (11)$$

where the x -axis is in the direction of the excited orbital, and $\beta = (m\omega_0/\hbar)^{1/2}$, $\hbar\omega_0$ being the energy separation between the states. From the wavefunctions the transition current density is immediately found, i.e.

$$\mathbf{j}_{1x}(\mathbf{r}) = \frac{e\hbar}{2im} \sqrt{2} \frac{\beta^4}{\pi^{3/2}} e^{-\beta^2 r^2} \mathbf{e}_x, \quad (12)$$

where \mathbf{e}_x is a unit vector in the direction of the excited orbital. In sec. 5 also the integrated current density \mathbf{J}_{1x} is needed. Using eq. (12) the explicit expression for this quantity becomes

$$\mathbf{J}_{1x} \equiv \int \mathbf{j}_{1x}(\mathbf{r}) d^3r = \frac{e\hbar}{im} \frac{\beta}{\sqrt{2}} \mathbf{e}_x. \quad (13)$$

4. Calculation of the local field

In general the Green's function can be split into two parts: a direct part (D) and an indirect part (I). In turn, the function $\mathbf{F}_{1x}(\mathbf{r})$ consists of two contributions

$$\mathbf{F}_{1x}(\mathbf{r}) = \mathbf{F}_{1x}^D(\mathbf{r}) + \mathbf{F}_{1x}^I(\mathbf{r}). \quad (14)$$

The indirect part corresponds to field propagation involving one or more reflections at the boundaries of the geometry in question. Consequently this part vanishes for a current source embedded in an infinite homogenous medium. The remaining direct Green's function $\mathbf{D}(\mathbf{q})$ is discussed in [7] and [10]. This part contains a singularity in Fourier space,

and so the integrations must be performed carefully. After a rather lengthy calculation it is found that

$$\mathbf{F}_{1x}(\mathbf{r}) = \frac{1}{2\pi^2} \frac{e\hbar}{im} \frac{\beta}{\sqrt{2}}$$

$$\begin{pmatrix} \left[\left(\frac{1}{r} - \frac{1}{q_0^2 r^3} \right) \Omega_1(r) + \frac{1}{(q_0 r)^2} \Omega_2(r) \right] \sin \gamma \\ 0 \\ \left[\left(\frac{1}{r} + \frac{2}{q_0^2 r^3} \right) \Omega_1(r) - \frac{2}{(q_0 r)^2} \Omega_2(r) - \frac{1}{q_0^2 r} \Omega_3(r) \right] \cos \gamma \end{pmatrix}. \quad (15)$$

Here $q_0 = (\omega/c)\sqrt{\epsilon(\omega)}$, $\epsilon(\omega)$ being the dielectric function of the ambient medium, γ is the angle between \mathbf{r} and \mathbf{e}_x , and the Ω functions are defined through

$$\Omega_n(r) = \int_0^\infty \frac{q^n}{q_0^2 - q^2} f_n(qr) e^{-q^2/4\beta^2} dq \quad (16)$$

with

$$f_n(qr) = \begin{cases} \cos qr & \text{for } n \text{ even} \\ \sin qr & \text{for } n \text{ odd} \end{cases}$$

As anticipated the integrals contain a singularity. Due to the generally complex values of $\epsilon(\omega)$, however, the poles $q = \pm q_0$ will never be located on the real q -axis. Consequently absolute convergence can be assumed. The explicit expressions for these integrals are obtained using results in [11]. The three components of the above vector are defined as follows: the (3) component is along \mathbf{r} , the (1) component is chosen so that \mathbf{e}_x is contained in the (1)-(3) plane, and finally the (2) component is orthogonal to the (1)-(3) plane. In the limit $r \rightarrow 0$ the "near-field" is obtained. Using the general expression above and the approximation $|q_0| \ll \beta$ it is found that

$$\mathbf{F}_{1x}(0) \approx \frac{\sqrt{2}}{6\pi^{3/2}} \frac{e\hbar}{im} \frac{\beta^4}{q_0^2} \mathbf{e}_x. \quad (17)$$

Consequently it is verified that no singularities exist in the near-field.

5. Local field factor

In order to find the numerical value of the local field the quantities γ_{1x}^0 and N need to be evaluated. As the background field \mathbf{E}^0 is assumed to be slowly varying one immediately has [cf. eq. (9)]

$$\gamma_{1x}^0 \approx \mathbf{E}^0(0) \cdot \mathbf{J}_{1x}. \quad (18)$$

The quantity N may in general be decomposed according to $N = N^D + N^I$. (19)

The indirect term vanishes in the present context, and the remaining term can be calculated using polar coordinates with γ as the polar angle. When the assumption $|q_0| \ll \beta$ is used, the following result is found

$$N = -\left(\frac{e\hbar}{m}\right)^2 \frac{\sqrt{2}\beta^5}{24\pi^{3/2}q_0^2}. \quad (20)$$

Using eqs. (6), (17), (18) and (20) the field at $r = 0$ can be written

$$E(0) = LE^0(0); \quad (21)$$

where the local field factor L , which here is defined as the ratio between the local field and the background field, is a scalar rather than a tensor. From the result obtained above it is found that

$$L = 1 - \frac{2\sqrt{2}aN}{1 + aN}. \quad (22)$$

Let us now apply the present theory to a GaAs quantum dot embedded in a $\text{Ga}_{1-x}\text{Al}_x\text{As}$ medium. For this system the dielectric function can be calculated using data from [12]. As electronic parameters the following values are used: $\hbar\omega_0 = 110 \text{ meV}$, $\hbar\nu = 7.5 \text{ meV}$ and $m = 0.068m_0$, where m_0 is the free electron mass. These values correspond to a wavefunction extent $\sim 1/\beta$ of approximately $1/\beta \approx 30 \text{ \AA}$. The quantity a is evaluated using the assumption that the Fermi-level lies in between the two levels. In the low temperature approximation one consequently has $\rho_{11}^F = 1$ and $\rho_{22}^F = 0$ so that

$$a = \frac{\mu_0}{\hbar} \frac{2\omega_0}{(\omega + i\nu)^2 - \omega_0^2}. \quad (23)$$

Using these results, L has been calculated as a function of ω . The result is shown in Fig. 1.

A significant feature of this curve is that the local field resonance is blue-shifted with respect to the electronic transition frequency. The shift can be found from the condition

$$\text{Re}[1 + aN] = 0. \quad (24)$$

In the frequency range of interest one can take $\varepsilon(\omega) \approx \varepsilon_\infty \approx 11.1$ [12]. The general solution to the resonance condition in the case where $\nu \approx 0$, then is given by

$$\omega = \omega_0 \left(\frac{1}{2} + \left(\frac{1}{4} + \frac{\Omega_0^4}{\omega_0^4} \right)^{1/2} \right)^{1/2}, \quad (25)$$

where

$$\Omega_0^4 = \frac{\sqrt{2}e^2\omega_0^3}{12\pi^{3/2}\varepsilon_0\varepsilon_\infty\hbar} \beta.$$

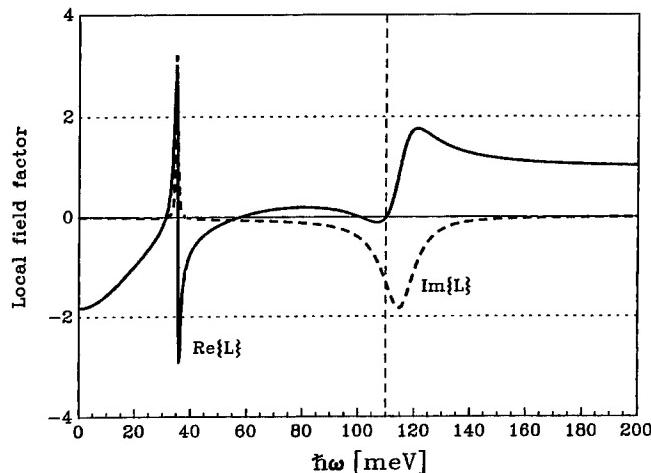


Fig. 1. The local field factor L as a function of ω . The vertical dashed line indicates the electronic transition frequency. The additional resonance around 35 meV is the phonon resonance in GaAs.

On the basis of eq. (25) the blue-shift Δ of the resonance becomes

$$\hbar\Delta = \hbar\omega - \hbar\omega_0 \approx 5.1 \text{ meV}. \quad (26)$$

It must be emphasized that the predicted shift has a different physical origin than the blue-shift, which is normally discussed in relation to quantum dots. The confinement of the electrons itself leads to a blue-shift of the transition frequency [1]. Thus, this shift is a result of a change in the electronic Hamiltonian (H^0) due to confinement, whereas the shift obtained in the present work is a result of a confinement-dependent change in the interaction Hamiltonian (H^{int}) between the electrons and the electromagnetic field.

In order to provide a qualitative demonstration of the results of the present treatment the normalized local field is plotted in Fig. 2 as a function of the radial distance from the center of the potential. Here the frequency is taken as $\hbar\omega = 120 \text{ meV}$, the background field is polarized along e_x , and the value $\gamma = \pi/2$ is used.

In [4] it is stated that for quantum dots of high symmetry the internal field will be uniform and slowly varying across the dot, except in the case of strong resonances. These arguments are based on a comparison between the wavelength of the optical field and the size of the quantum dot. The present work shows that this is an inadequate measure of the effect of nonlocality, as it is the spatial behavior of the local field and not the background field, which determines the "degree of nonlocality".

6. Summary

In this paper we have derived a general expression for the local field inside a quantum dot embedded in a homogenous medium. Starting from a self-consistent integral equation, the direct part of the local field is identified. The electronic confinement in the quantum dot is modelled by an isotropic, parabolic potential, and only transitions between the ground state and the first excited state are taken into account. It is verified that the magnitude of the local field has a finite value at all positions in space. Finally a resonance condition is set up for the near-field, and it is shown that the local field resonance frequency is blue-shifted with

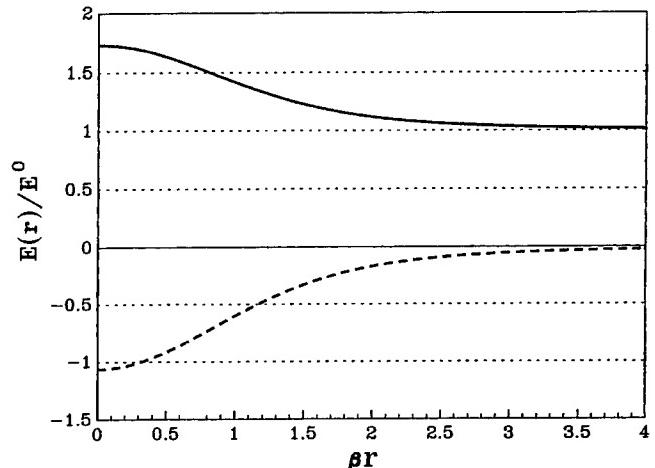


Fig. 2. The normalized local field as a function of the radial distance r . The real and imaginary parts are shown as the solid line and the dashed line, respectively. The frequency is taken as $\hbar\omega = 120 \text{ meV}$.

respect to the electronic transition frequency. For a GaAs/Ga_{1-x}Al_xAs quantum dot with a wavefunction extent of 30 Å and a transition frequency of $\hbar\omega_0 = 110$ meV a shift of 5.1 meV is predicted. The calculation of the local field shows that the internal field inside the quantum dot may have a strong spatial dependence, even when the background field is slowly varying. Consequently an important conclusion is that nonlocal effects are of significant importance also in the case, where the optical wavelength is much larger than the extent of the quantum dot. In the present treatment the diamagnetic contribution to the current density is neglected, thus violating charge conservation. Consequently the validity of the results is restricted to the region $\omega \approx \omega_0$, where the paramagnetic contribution dominates. The inclusion of diamagnetic effects is expected to be important outside this region but the general conclusions regarding the significance of nonlocality are believed to be valid still. The effect of diamagnetism on the blue-shift of the local field resonance is still unclear.

References

1. Brunner, K. *et al.*, Phys. Rev. Lett. **69**, 3216 (1992).
2. Sikorski, Ch. and Merkt, U., Phys. Rev. Lett. **62**, 2164 (1989).
3. Gaponenko, S. V. *et al.*, J. Opt. Soc. Amer. B**10**, 1947 (1993).
4. Schmitt-Rink, S., Miller, D. A. B. and Chemla, D. S., Phys. Rev. B**35**, 8113 (1987).
5. Leonelli, R. *et al.*, Phys. Rev. B**48**, 11135 (1993).
6. Takagahara, T., Phys. Rev. B**47**, 16639 (1993).
7. Keller, O., Physics Report (to be published).
8. Keller, O., in: "Studies in Classical and Quantum Nonlinear Optics" (Edited by O. Keller) (Nova Sciences, New York 1994).
9. Feibelman, P. J., Progr. Surf. Sci. **12**, 287 (1982).
10. Keller, O., J. Opt. Soc. Amer. B (in press).
11. Gradshteyn, I. S. and Ryznik, I. M., "Table of Integrals, Series and Products" (Academic, London 1980).
12. Eliasson, G., Hawrylak, P. and Quinn, J. J., Phys. Rev. B**35**, 5569 (1987).

Resonant Tunneling: From Model Hamiltonian to Modern Electronic Devices

K. A. Chao

Division of Physics, Faculty of Physics and Mathematics, Norwegian Institute of Technology, University of Trondheim, N-7034 Trondheim, Norway

M. Willander

Department of Physics and Measurement Technology, University of Linköping, S-581 83, Linköping, Sweden

and

Yu. M. Galperin

Department of Physics, University of Oslo, P.O. Box 1048 Blindern, N-0316 Oslo 3, Norway, and A. F. Ioffe Physico-Technical Institute, 194021 St. Petersburg, Russia

Received June 6, 1994; accepted June 15, 1994

Abstract

The single-electron theory of resonant-tunneling through a double-barrier structure with perfect interfaces has been reviewed, and the future direction of theoretical development including the electron-electron interaction and rough interfaces is outlined. The theoretical understanding of this system reveals its potential application to quantum electronic devices.

1. Introduction

The molecular-beam epitaxy (MBE) technique has opened an entirely new area of condensed matter physics including materials fabrication, experimental characterization, theoretical investigation, and device designs. So far the most popular field of research is based on the MBE-fabricated low-dimensional III-V systems such as quantum dots, quantum wires, quantum wells and superlattices. Because of the high carrier mobility in these heterostructures, they are potential systems for modern high speed electronic devices. However, one important aspect of the III-V semiconductors is their far superior optical properties over the group IV semiconductors. It is therefore of great interest to combine the transport and optical properties of III-V materials, in an effort towards the creation of a new generation of quantum devices.

In this paper we will restrict ourselves to the double-carrier (DB) resonant-tunneling (RT) systems which, under a DC bias V is schematically illustrated in Fig. 1, where interfaces are parallel to the xy -plane and the tunneling is along the z -axis. Ignoring the external circuit, from the left to the right, the DBRT structure consists of a heavy doped region, a spacer (S), a barrier, a well, another barrier, another spacer (S), and another heavy doped region. In a bipolar DBRT structure, the doping is n-type at one end, but p-type at the other end. In this case one has to consider both the electron tunneling through the conduction band, and the hole tunneling through the valence band, as indicated by the upper diagram in Fig. 1. A simpler system is the unipolar DBRT structure in which the dopings are either n-type or p-type. Hence, only one energy is involved in the tunneling transport. For simplicity, let us consider the n-doped unipolar

DBRT structure shown in the lower diagram in Fig. 1. Depending on the structure of a DBRT diode, there may exist bound states in the spacer at the emitter side. If electrons tunnel from the emitter Fermi sea (or the bound state), we call the emitter a 3D (or 2D) emitter.

In this paper we will first review the important theoretical works on tunneling through a n-doped unipolar DBRT structure, within the framework of single-electron approximation and assuming perfect interfaces. In most general form the total Hamiltonian $\mathcal{H} = \mathcal{H}_{el} + \mathcal{H}_f + \mathcal{H}_s$, consists

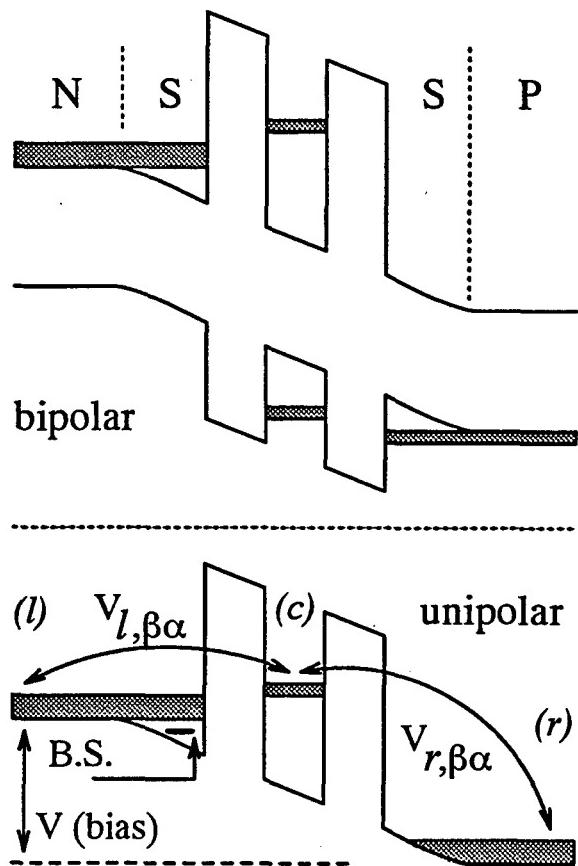


Fig. 1. Schematic illustration of a bipolar and a unipolar double-barrier resonant-tunneling structure.

of an electronic part

$$\mathcal{H}_{el} = \sum_{j, \beta} \varepsilon_{j, \beta} c_{j, \beta}^\dagger c_{j, \beta} + \sum_\alpha E_\alpha c_\alpha^\dagger c_\alpha + \sum_{j, \alpha, \beta} [V_{j, \beta \alpha} c_\alpha^\dagger c_{j, \beta} + h.c.], \quad (1)$$

where $j = l$ (or r) refers to the left emitter (or right collector), a part \mathcal{H}_f for the surrounding field, and a part \mathcal{H}_s representing the interaction between the electron and the field. $\varepsilon_{j, \beta}$ is the electronic energy in the emitter or the collector, and E_α is the quasibound state energy in the well with width Γ . The tunneling matrix elements $V_{j, \beta \alpha}$ are calculated according to Bardeen's prescription [1] using the potential in Fig. 1. In this review we will consider \mathcal{H}_f for the phonon field, the external uniform magnetic field, and an AC bias.

2. Single-electron theory of RT with perfect interfaces

The first case we consider is the electron–phonon interaction which breaks the phase coherence. The observed phonon-replica in the I – V characteristic² was previously analyzed by several theoretical attempts^{3–7} using a one dimensional phonon model. This model has neglected the electron recoils when it scatters inelastically against a LO phonon and therefore underestimates the electron–phonon interaction.

Recently, a three-dimensional model was considered [8], and the results so-obtained will be outlined here. The electronic energies in a 3D emitter and collector are $\varepsilon_{j, p} = \varepsilon_{j, p} + \varepsilon_{k\parallel}$, where p is the momentum component along the z -axis, and $\varepsilon_{k\parallel} = \hbar^2 k_\parallel^2 / 2m^*$. For simplicity, only one quasi-bound state ($\alpha \equiv c$) energy $E_c = \varepsilon_c$ is considered. The phonon Hamiltonian $\mathcal{H}_f \equiv \mathcal{H}_{ph}$ is simply

$$\mathcal{H}_{ph} = \sum_{i, \mathbf{Q}, q} \hbar \omega_i(\mathbf{Q}, q) b_{i, \mathbf{Q}, q}^\dagger b_{i, \mathbf{Q}, q}, \quad (2)$$

where q is the phonon momentum along the z -axis. The inelastic electron–LO phonon scatterings are expressed in terms of the Fröhlich Hamiltonian [9]

$$\mathcal{H}_s = \sum_{i, \mathbf{Q}, q, \mathbf{k}\parallel} M_i(\mathbf{Q}, q) (b_{i, -\mathbf{Q}, q}^\dagger + b_{i, \mathbf{Q}, q}) c_{c, \mathbf{k}\parallel + \mathbf{Q}}^\dagger c_{c, \mathbf{k}\parallel}. \quad (3)$$

The essential theoretical work is to calculate the probability $T(\varepsilon_\perp, \mathbf{k}\parallel)$ for an electron occupying the state $(\varepsilon_\perp, \mathbf{k}\parallel)$ in the left emitter to be transmitted to the right collector, from which the tunneling current can be obtained. $T(\varepsilon_\perp, \mathbf{k}\parallel)$ can be formulated according to the scattering theory, and the wide band approximation³ (WBA) $\sum_\beta |V_{j, \beta, \alpha=c}|^2 \delta(\varepsilon_\perp - \varepsilon_{j, \beta}) \simeq \Gamma_j$ has been commonly used. The WBA is a justified approximation, but special care should be taken if the energy ε_c is very close to the bottom of the conduction band edge in the emitter. After applying the WBA, the simplified form of $T(\varepsilon_\perp, \mathbf{k}\parallel)$ is then analyzed with a cumulant expansion, which can easily include higher order vertex contributions. It is important to mention that the weak LO phonon dispersion is usually neglected.

It was found in Ref. [8] that the I – V characteristics depends strongly on the Fermi energy (or the doping concentration) in the emitter. When an electron in the emitter with total energy $\varepsilon_\perp + \varepsilon_{k\parallel}$ tunnels into the well and emits a LO phonon, the final energy of the electron $\varepsilon'_\perp + \varepsilon_{k\parallel}$

satisfies the conservation of energy $\varepsilon'_\perp = \varepsilon_\perp - \hbar\omega(\mathbf{Q}, q) + \varepsilon_{k\parallel} - \varepsilon_{k\parallel}'$, where $\mathbf{k}'_\parallel = \mathbf{k}_\parallel - \mathbf{Q}$. Therefore, if the Fermi energy is comparable or larger than the LO phonon energy, we may have $\varepsilon'_\perp \simeq \varepsilon_\perp$. Consequently, as the Fermi energy in the emitter is increased, the electron-recoil broadens the phonon-replica, which is also shifted towards the main peak. When the Fermi energy reaches the LO-phonon energy, the phonon-replica and the main peak merge into one. This mechanism can change drastically the peak-to-valley ratio of the tunneling current.

The second case we review is the RT through a DB structure under a uniform magnetic field \mathbf{B} . There are two interesting cases: $\mathbf{B} \parallel \mathbf{I}$ and $\mathbf{B} \perp \mathbf{I}$. For $\mathbf{B} \perp \mathbf{I}$, the Lorentz force modifies \mathbf{k}_\parallel as the electron tunnels through the barrier at the emitter side. Therefore, the field-free RT condition will be destroyed, but can be restored by adjusting the bias. In this way the electronic dispersion relation $\varepsilon_{k\parallel}$ in the quantum well can be measured.

The more interesting situation is $\mathbf{B} \parallel \mathbf{I}$. There have been extensive experimental effort to deduce from the I – V curve and the I – B curve the effective mass, the charge buildup in the well, and the dimensionality (2D or 3D) of the emitter (see references in Ref. [10]). However, here we will outline a recent theoretical study on phonon-assisted resonant magnetotunneling (PARMT) through a DBRT diode [10]. The results are more interesting if the emitter is 2D, for which all electronic energies in (1) are simply Landau levels $\varepsilon_j + (n_j + \frac{1}{2})\hbar\omega_c^l$. Here we assume only one bound state energy ε_c in the emitter spacer ($j = l$), the well ($j = c$), and the collector spacer ($j = r$), and ω_c^l is the corresponding cyclotron frequency. To describe unambiguously the important results, let us consider a dispersionless phonon spectrum with energy $\hbar\omega_0$. At low temperature one expects the formation of magnetic polaron due to the phonon emission, which was indeed observed [11].

At finite temperature there is also phonon absorption process which has never been observed in bulk semiconductors, but may be detected in a DBRT diode because of the spatial confinement of the wave functions. Without inelastic electron–phonon scattering, the RT obeys the selection rule $\varepsilon_l + (n_l + \frac{1}{2})\hbar\omega_c^l = \varepsilon_c + (n_l + \frac{1}{2})\hbar\omega_c^c$. However, the condition for PARMT is $\varepsilon_l + (n_l + \frac{1}{2})\hbar\omega_c^l = \varepsilon_c + (n_c + \frac{1}{2})\hbar\omega_c^c - v\hbar\omega_0$, where v is a positive integer. Consequently, if $(n_c - n_l)\hbar\omega_c^c = v\hbar\omega_0$, then the direct RT and the PARMT occur at the same bias voltage. This double resonance has a single broad peak in the I – V curve. On the other hand, if $(n_c - n_l \pm \frac{1}{2})\hbar\omega_c^c = v\hbar\omega_0$, the direct RT current peak is very sharp and is sandwiched between two satellite PARMT peaks. The theoretical analysis of PARMT follows the similar procedure as the case discussed above without an external magnetic field. In reality, $v = 1$ with available high magnetic field. The details study with numerical calculation [10] discovered a characteristic oscillation of the width and height of the main peak in the I – V curve when the ratio ω_0/ω_c^c changes between integer and half-integer. Such phenomenon exhibits even at a temperature of 200 K. Therefore, if confirmed experimentally, this theoretical prediction provides an additional method to determine the electronic effective mass (via ω_c^c) in the quantum well under the dynamical RT condition.

Finally we review the RT under a combined DC–AC bias. In this case the total Hamiltonian is simply eq. (1) with the

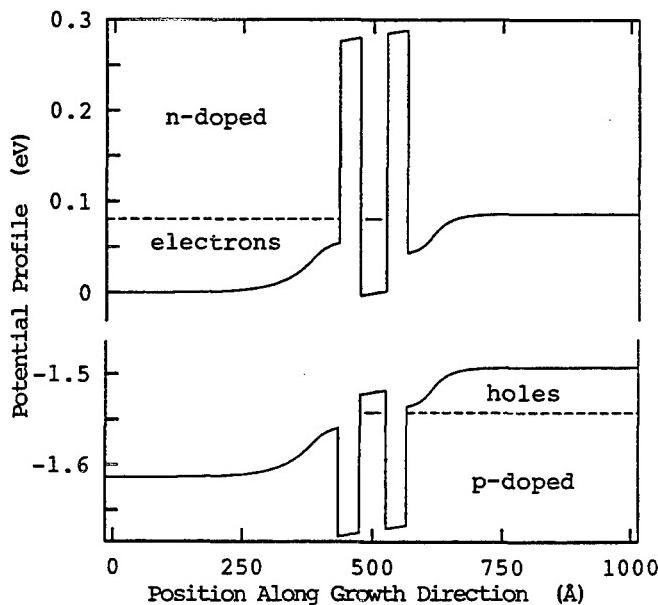


Fig. 2. Selfconsistent potential profile and energy levels of a bipolar RTLED under the condition of simultaneous RT. The dotted lines marks the local Fermi energies of electrons and holes.

electronic energy $\epsilon_{j,\beta} = \epsilon_{j,p}(t) + \epsilon_{k\parallel}$, and $E_\alpha = \epsilon_c(t) + \epsilon_{k\parallel}$. This Hamiltonian has been studied by several authors [12–18]. However, the central theme of the problem is the characteristic response of a DBRT diode to an applied bias. Is the response inductive or capacitive? Since the quasibound state in the well has a finite life time $1/\Gamma$, there is a time delay of the current following the switch-on of the bias, resulting in an intrinsic inductive characteristics of each RT process. Only very recently a thorough theoretical analysis [19] has cleared this physical picture, and provided a solid framework for high frequency devices applications. With a combined DC-AC bias $V_0 + V(\omega t) = V_0 + V_1 \sin \omega t$, the current $I(\omega, t) = I_0(\omega) + I_1(\omega, t)$ can be solved and so the admittance $Y(\omega)$ is derived, from which we obtain the AC conductance $G(\omega) \equiv [\text{Re } Y^{-1}(\omega)]^{-1}$ and the intrinsic inductance $L(\omega) \equiv 1/\omega \text{Im } Y^{-1}(\omega)$.

Knowing the conductance and the inductance, we can construct a quantum equivalent circuit of the DBRT diode. From this equivalent circuit the upper limit of operating frequency ω_{max} for the DBRT diode can be calculated rather accurately. As an example, using the diode of Brown *et al.* [20] the theoretical value is $\omega_{max} = 187 \text{ GHz}$ for the temperature 300 K , in good agreement with the observed $\omega_{max} \approx 200 \text{ GHz}$.

3. Further complications

Our review in the previous section has neglected two important aspects: the interface-roughness scattering (SRS) and the electron-electron interaction (EEI) in the well. The SRS is the dominating inelastic scattering channel if the DBRT structure has thick and low barriers. Works in this area were summarized in a recent paper by Johansson [21], where he has shown a reduction of about 10% in peak current, but a several order of magnitude increase in valley current by SRS. Such dramatic decrease of peak-to-valley ratio is of crucial importance for quantum device designs.

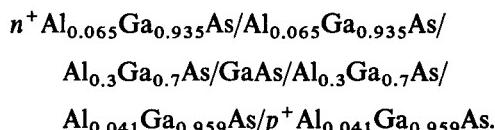
For a realistic DB diode under RT condition, the density of charge accumulated in the well is of the order 10^{11} cm^{-2} .

Via the EEI, this accumulated charge modifies the effective single-electron potential and so alters all parameters in the Hamiltonian. A very important direct consequence of the EEI is the observed intrinsic bistability in the $I-V$ curve [22]. Earlier works on this problem included only the Coulomb term of the EEI [23, 24]. A recent study [25] including the exchange-correlation term within the random phase approximation (RPA) indeed improved the agreement with experiments. However, the validity of RPA for the charge density in question is doubtful.

It is obvious that for DBRT structures, at least two issues need to be clarified in the future: the combined effect of electron-phonon interaction and surface-roughness scattering, and a better treatment of the exchange-correlation effect than the simple RPA.

4. Selfconsistent scheme for device modelling

As we mentioned in the Introduction, the potential application of DBRT diodes made from III-V semiconductor lies in optical devices, but making use of the fast RT to bring electrons and holes into the well. Here we consider a RT light emitting diode (LED), the main part of which is a bipolar DBRT structure. For such device modelling, one needs to solve selfconsistently the Schrödinger equation and the Poisson equation. The selfconsistent modelling suggests the following RTLED. The symmetric bipolar DBRT diode has the following specific structure with 50 Å spacers.



The doping concentration is $2 \times 10^{18} \text{ cm}^{-2}$ for donor and $4 \times 10^{18} \text{ cm}^{-2}$ for acceptor. Under a forward bias of -1.6 Volts the DBRT diode achieves simultaneous RT of both electrons and holes, as shown in Fig. 2. The in-tunneling electrons and holes are confined in the well to recombine to emit light. This is the first step to make an efficient LED.

Acknowledgements

This work was supported by Nordisk Forskerutdanningsakademi via a Research Network on Physics of Nanometer Electronic Devices.

References

1. Bardeen, J., Phys. Rev. Lett. **6**, 57 (1961).
2. Goldman, V. J., Tsui, D. C. and Cunningham, J. E., Phys. Rev. **B36**, 7635 (1987).
3. Wingreen, N. S., Jacobsen, K. W. and Wilkins, J. W., Phys. Rev. Lett. **61**, 1396 (1988).
4. Glazman, L. I. and Shekhter, R. I., Zh. Eksp. teor. Fiz. **94**, 292 (1988) [Sov. Phys. JETP **67**, 163 (1988)].
5. Cai, W., Zheng, T. F., Hu, P., Yudanin, B. and Lax, M., Phys. Rev. Lett. **63**, 418 (1989).
6. Jonson, M., Phys. Rev. **B39**, 5924 (1989).
7. Lake, R. and Datta, S., Phys. Rev. **B45**, 6670 (1992), and references therein; Superlatt. Microstruc. **11**, 83 (1992).
8. Zou, N. and Chao, K. A., Phys. Rev. Lett. **69**, 3224 (1992).
9. Mori, N. and Ando, T., Phys. Rev. **B40**, 6175 (1989).
10. Zou, N., Chao, K. A. and Galperin, Yu. M., Phys. Rev. Lett. **71**, 1756 (1993).
11. Boebinger, G. S., Levi, A. F. J., Schmitt-Rink, S., Passner, A., Pfeiffer, L. N. and West, K. W. Phys. Rev. Lett. **65**, 235 (1990).
12. Coon, D. D. and Liu, H. C., Appl. Phys. Lett. **49**, 94 (1986).
13. Frenslay, W. R., Appl. Phys. Lett. **51**, 448 (1987).

14. Wingreen, N. S., *Appl. Phys. Lett.* **56**, 253 (1990).
15. Jacoboni, C. and Price, P. J., *Solid State Commun.* **75**, 193 (1990).
16. Liu, H. C., *Phys. Rev. B* **43**, 12538 (1991).
17. Chen, L. Y., and Ting, C. S., *Phys. Rev. Lett.* **64**, 3159 (1990).
18. Johansson, P., *Phys. Rev. B* **41**, 9892 (1990).
19. Zou, N., Willander, M. and Chao, K. A., *Phys. Rev. B – Rapid Commun.* (in press).
20. Brown, E. R., Parker, C. D. and Sollner, T. C. L. G., *Appl. Phys. Lett.* **54**, 934 (1989).
21. Johansson, P., *Phys. Rev. B* **48**, 8938 (1993).
22. For a review, see L. Eaves, F. W. Sheard and G. A. Toombs, "Physics of Quantum Electron Devices" (Edited by F. Capasso) (Springer, Berlin 1990).
23. Sheard, F. W. and Toombs, G. A., *Appl. Phys. Lett.* **52**, 1228 (1988).
24. Rahman, M. and Davies, J. H., *Semicond. Sci. Technol.* **5**, 168 (1990).
25. Zou, N., Willander, M., Linnerud, I., Hanke, U., Chao, K. A. and Galperin, Yu. M., *Phys. Rev. B* **49**, 2193 (1994).

Aharonov-Bohm Effect in a Quantum Ring with Strong Electron-Electron Correlations

I. V. Krive,^{1,2} R. I. Shekhter,¹ S. M. Girvin³ and M. Jonson¹

¹ Department of Applied Physics, Chalmers University of Technology and Göteborg University, S-412 96 Göteborg, Sweden

² B. Verkin Institute for Low Temperature Physics and Engineering, Academy of Sciences of Ukraine, 310164 Kharkov, Ukraine

³ Department of Physics, Indiana University, Bloomington, Indiana 47405, U.S.A.

Received June 28, 1994; accepted June 28, 1994

Abstract

We calculate the magnetic moment ("persistent current") in a strongly correlated electron system – a Wigner crystal – in a one-dimensional ballistic ring (quantum ring). The flux- and temperature dependence of the persistent current is shown to be essentially the same as for a system of non-interacting electrons. In contrast, by incorporating into the ring geometry a tunnel barrier, that pins the Wigner crystal, the current is suppressed and its temperature dependence is drastically changed. The competition between two temperature effects – a reduced barrier height for macroscopic tunneling and a loss of quantum coherence – results in a sharp peak in the temperature dependence, which for a rigid Wigner crystal appears at $T \sim 0.5\hbar s/L$, (s is the sound velocity of the Wigner crystal, L is the length of the ring).

In recent experiments [1, 2] persistent currents have been observed in the ballistic transport regime of mesoscopic rings formed in the laterally confined two-dimensional electron gas of certain AlGaAs heterostructures [3]. The current I and the associated magnetic moment were found to oscillate as a function of magnetic flux with period $\Phi_0 = hc/e$ – the quantum unit of flux – and amplitude $I_0 \sim ev_F/L$ (e is the electronic charge, v_F the Fermi velocity, and L the length of the circumference of the ring). These results are in excellent agreement with a theory of such Aharonov-Bohm (AB) oscillations based on a free electron model of the ballistic electrons [4, 5]. Since electron-electron interactions in the semiconductor ring are not weak, and since electron correlations must play an important role when the density of conducting electrons is low, this agreement is quite surprising. In diffusive metal rings, for example, where the electronic mean free path is short ($l \ll L$), it has been suggested that electron correlations significantly enhance the amplitude of the AB oscillations [6]. Thus the question of how Coulomb correlations in a system of ballistic electrons affect the magnitude of the persistent current is of significant interest.

In this letter we study persistent currents and AB oscillations in systems of spinless interacting electrons confined to a one-dimensional ring; the electrons are assumed to be so strongly correlated that they form a Wigner crystal. In an ideal ring the mechanism of the persistent current is a dissipationless sliding of the crystal as a whole. We demonstrate that the resulting current oscillates as a function of magnetic flux with period Φ_0 . Its amplitude at low temperatures is exactly the same, $I_0 = ev_F/L$ ($v_F = \pi\hbar/m a$ is the Fermi velocity, a is the period of the Wigner crystal, m is the electron mass) as for noninteracting electrons of the same density in accordance with a general theorem [7]. If the

temperature is raised, the amplitude of the oscillations is exponentially suppressed: $I(T) \sim I_0 \exp(-\pi T/2T_0)$, where $T_0 \equiv \hbar v_F/L$ is the characteristic crossover temperature. Therefore the magnitude as well as the temperature dependence of the persistent current carried by an ideal Wigner crystal looks completely identical to that of a current carried by a free electron gas.

The situation changes drastically if a potential barrier, somewhere along the ring, impedes the motion of the electrons. Charge transport in this case requires that electrons tunnel through the barrier – the process which for strongly correlated electrons can be viewed as a macroscopic tunneling of a Wigner crystal-ring. In the case of high enough barrier (strong pinning) it is convenient to think of the motion of the crystal as a two-step process, where first a single electron tunnels through the barrier producing a deformation of a finite portion of the Wigner crystal, which then is relaxed [8, 9]. This process necessarily depends on the elastic properties of the crystal, and as a result the magnitude of the persistent current will depend on the sound velocity s in the Wigner lattice. As our analysis below will show, the temperature dependence of the amplitude of the AB oscillations is also affected in a qualitative way. The presence of the tunnel barrier, which pins the Wigner crystal and makes charge transfer possible only by macroscopic tunneling, strongly decreases the zero temperature value of the persistent current since for a repulsive interaction quantum fluctuations in a strongly correlated electron system renormalize the barrier upward. The finite ring circumference cuts off the divergent renormalization of the barrier height which occurs in the thermodynamic limit of a Luttinger liquid [10] or Wigner crystal [8, 9]. Thus the persistent current at zero temperature is greatly reduced but is still finite. The competition between two effects of an increased temperature – a temperature stimulated tunneling and a loss of phase coherence due to the enhancement of destructive interference – leads to a sharp maximum in the temperature dependence of the persistent current. For a rigid crystal this maximum occurs at $T \sim 0.5T_s$, where $T_s \equiv \hbar s/L \gg T_0$. This effect makes it possible to measure the Wigner crystal sound velocity in a ring with an "adjustable barrier" (height controlled by a gate voltage).

The starting point of our analysis is a model system, where the Wigner crystal is regarded as an elastic chain of spinless electrons forming a ring. In the presence of a potential barrier, smooth on the scale of a but well localized on the scale of L , the Lagrangian of such a system in the long

wavelength approximation is [9]

$$L = \frac{ma}{8\pi^2} \{ \dot{\phi}^2 - s^2(\phi')^2 \} - V_0 \delta(x) \cos(\phi). \quad (1)$$

Here $\phi = 2\pi u(x)/a$ is the dynamical displacement field of the crystal and V_0 is the magnitude of the pinning potential (without loss of generality placed at the point $x = 0$).

We emphasize that (1) is an effective Lagrangian that describes the long wavelength aspects of the quantum dynamics of the Wigner crystal. The short wavelength fluctuations do not affect the global dynamics of the system, and only result in a renormalization of the magnitude V_0 of the potential, (already included in (1) but negligible for a stiff Wigner crystal [9]). We assume that the ring circumference is large enough to justify dropping terms from (1) which are irrelevant in an infinite system.

In the presence of a magnetic field, directed normal to the plane of the ring, the one-dimensional Lagrangian (1) acquires an additional term, L_{int} . This term describes the AB interaction of the Wigner crystal with the vector potential of an electromagnetic field, $A_\phi = \Phi/L$ (Φ is the magnetic flux through the ring). The AB interaction term, rewritten using the real scalar displacement field ϕ , has the form of a total time derivative,

$$L_{int} = (\hbar/L)(\Phi/\Phi_0)\dot{\phi}, \quad (2)$$

and affects, as must be the case, only the quantum dynamics of the crystal.

The flux-induced persistent current $I(\Phi) = -c\partial F/\partial\Phi$, is defined in terms of the sensitivity of the free energy of the ring to a magnetic flux. For the following analysis, it is convenient to express the free energy F as a functional integral over quantum- and thermal fluctuations of the displacement field,

$$F = -k_B T \ln \left\{ \sum_{n=-\infty}^{\infty} (-1)^{n(N-1)} \int D\phi_n e^{-S_E[\phi_n]/\hbar} \right\}, \quad (3)$$

where the action S_E derives from the Lagrangian (1), (2) in the imaginary time representation. "Twisted" boundary conditions in imaginary time are imposed on the field ϕ (see, e.g. [11]):

$$\phi_n(\tau + \beta, x) = \phi_n(\tau, x) + 2\pi n. \quad (4)$$

Here $n = 0, \pm 1, \pm 2, \dots$ is the topological (winding) number, classifying homotopically inequivalent trajectories. The physical meaning of the boundary condition (4) follows from the definition of the field $\phi = 2\pi u(x)/a$; a uniform shift of the crystal by a distance equal to an integer times the lattice constant a leads, in the ring geometry, to a state identical to the initial state after certain permutations of electrons. For the minimum shift by $1 \times a$ ($\Delta\phi = 2\pi$), the initial state is recovered after $(N-1)$ successive permutations of pairs of electrons. The corresponding extra phase $\pi(N-1)$, that appears in the many-particle wave-function because the electrons obey Fermi statistics, generates the factor $(-1)^{n(N-1)}$ in (3). As we will see below, this factor properly accounts for the parity effects in the response of one-dimensional interacting electrons to a magnetic field [12–14]. We note in passing that the analogous twisted boundary conditions appear when the Luttinger model is applied to a ring geometry [13]. The appearance of the homotopy index n in the boundary condition (4) suggests

that the functional integral should first be calculated for trajectories belonging to a definite homotopy class, and then the homotopically non-equivalent classes of trajectories should be summed over.

In every homotopy class we will calculate the functional integral using the saddle point approximation, assuming the saddle point action to be large, $S_n \gg \hbar$, on the extremal trajectory given by the solution of the classical equations of motion in imaginary time. Below we will show that this assumption is justified for a stiff Wigner crystal ($\alpha \ll 1$).

First we calculate the persistent current in the ideal, unpinned crystal ($V_0 = 0$). In a perfect (or weakly pinned) Wigner crystal, long-wave quantum fluctuations are cut off at the wavelength of the order of the crystal size L . It is physically evident that we can imagine an ordered crystal structure as long as the mean square fluctuations of the dimensionless field ϕ ,

$$\langle \phi^2 \rangle \sim \alpha \int_{\pi/L}^{\pi/a} \frac{dk}{k} \coth \left(\frac{s\hbar}{2k_B T} k \right), \quad (5)$$

are small so that $\langle \phi^2 \rangle \ll 1$ (T is the temperature, $\alpha = \pi\hbar/msa$ is a dimensionless parameter that characterizes the strength of quantum fluctuations in the Wigner crystal). For $T \rightarrow 0$ this restriction imposes an upper bound on the chain length $L \ll ae^{1/\alpha}$; for such samples the thermal fluctuations are suppressed up to a temperature $T \lesssim T_s/\alpha$ ($T_s \equiv \hbar s/L$). The situation is changed drastically for a strongly pinned Wigner crystal where an "intermediate" cut off scale appears [9].

One can readily calculate the persistent current of an ideal ring as the problem in the long wavelength limit is described by a quadratic Lagrangian. The external trajectory corresponding to the boundary condition (4) is linear in imaginary time and independent of the x -coordinate,

$$\phi_n(\tau) = 2\pi n(\tau/\hbar\beta). \quad (6)$$

By substituting (6) into (1, 2, 3), it is easy to find an exact solution for the free energy in terms of the Jacobi function θ_3 (see e.g. [15]). The asymptotic expressions for the persistent current at high- and low-temperatures are

$$\frac{I_{WC}}{I_0} \simeq \begin{cases} 2 \frac{T}{T_0} e^{-(\pi/2)(T/T_0)} (-1)^N \sin \left(2\pi \frac{\Phi}{\Phi_0} \right), & T \gtrsim T_0 \\ 1 - 2 \left\{ \left\{ \frac{\Phi}{\Phi_0} + \delta_N \right\} \right\}, & T \ll T_0 \end{cases} \quad (7)$$

Here $\{x\}$ denotes the fractional part of x , and the parity dependent term δ_N is 1/2 (0) for N odd (even). Thus the persistent current carried by an ideal Wigner crystal is a periodic function of flux with period $\Phi_0 = hc/e$ and amplitude $I_0 = ev_F/L$ at low temperatures. The oscillations are exponentially damped at $T \gtrsim T_0 = \hbar v_F/L$. The current has a paramagnetic character when there is an even number of electrons in the ring (i.e. the induced magnetic moment is parallel to the external magnetic field) and diamagnetic for an odd number of electrons. All these properties of the persistent current coincide with those calculated using the model of an ideal Fermi gas. For $T = 0$, this was first shown in Ref. [7] for a general case of arbitrary Coulomb-like interaction.

At finite temperatures there are in general additional contributions due to crystal deformations produced by ther-

mally excited plasmons and due to electron-phonon interaction. It is possible to show [16] that in a perfect Wigner crystal ring the contribution of plasmon fluctuations to the action leads to the change of exponent in eq. (7) $T/T_0 \rightarrow (T/T_0)[1 + \alpha^3(T/E_F)^2]$ which is irrelevant if the temperature is less than the Fermi energy $E_F = mv_F^2/2$. As for inelastic electron-phonon scattering, it results in a “broadening”, $\hbar\gamma$, of the quantized energy levels of the quantum ring. This effect can be accounted for phenomenologically by assuming a dissipative character of the Wigner crystal dynamics. At temperatures $T \gg \hbar\gamma$ the dissipative corrections to the Euclidean action are of the form $\delta S_E^{(n)} \sim \hbar^2\gamma|n|/T_0$, which leads to an additional exponential suppression of the persistent current.

The thermal destruction of the persistent current can be characterized by a crossover temperature T_c , where $I \propto \exp(-T/T_c)$. From (7) one has $T_c = (2/\pi)(\hbar v_F/L)$, which is twice as large as the crossover temperature found for a ring of free electrons characterized by a constant chemical potential [5, 17]. Rather than with the electron-electron correlations [13] the factor of 2 difference is connected with the fact that in our case the number of electrons – not the chemical potential – is fixed [18].

Let us now consider the persistent current in a Wigner crystal in the presence of a potential barrier. A uniform sliding motion of the crystal is impossible in this case, and charge transport along the ring is connected with macroscopic quantum tunneling (MQT) of the Wigner crystal. The character of the MQT is dictated by the pinning strength. At strong pinning, $\alpha V_0 \gg T_s$, the mechanism for charge transport around the ring includes tunneling of a finite segment of the Wigner crystal through the barrier, as well as the subsequent relaxation of the associated elastically deformed state of the crystal. In the weak pinning regime, $\alpha V_0 \ll T_s$, the Wigner crystal as a whole tunnels through the barrier (without essential distortions). The above mechanisms for macroscopic tunneling were first considered in connection with the tunneling of commensurate charge density waves [8] and have also been used to describe the tunneling conductivity of a Wigner crystal [9]. In these contexts it was shown [8, 9] that in the case of strong pinning the dominating tunneling process is the elastic relaxation of the deformed state arising in the near-barrier region.

In the ring geometry a shift of the crystal as a whole by the lattice period, a , may include one or several “tunneling steps” (by a “tunneling step” we understand the combined processes of tunneling and relaxation of the elastic deformation). In the case of strong pinning, the single-step tunneling is described by the exact solution of the free equation of motion ($V_0 = 0$) in imaginary time with the twisted boundary condition (4) and $n = \pm 1$:

$$\phi_\sigma^{(s)}(\tau) = 2\sigma \arctan \left[\coth \left(\frac{\pi|x|}{\hbar s \beta} \right) \tan \left(\pi \left(\frac{\tau}{\hbar \beta} - \frac{1}{2} \right) \right) \right], \quad (8)$$

$$\sigma = \pm 1.$$

A description of the dominating relaxation process in terms of this “periodic instanton”-solution [19] is valid in the region outside the interval $[-l_0, l_0]$ containing the part of the crystal deformed by the initial tunneling process. The length l_0 , which is inversely proportional to the potential V_0 , appears only as a limit of the integration over coordi-

nate x ; we assume that $l_0 \ll L/2$, a criterion which one can show to be equivalent to a restriction on temperature, $T \ll \alpha V_0$.

A multi-step solution is a sequence of single steps of type (8), corresponding to all possible intermediate rotations of the crystal

$$\phi_\sigma^{(m)}(\tau) = \sum_i \phi_{\sigma_i}^{(s)}(\tau - \tau_i), \quad \sigma = \sum_i \sigma_i = \pm 1. \quad (9)$$

The set of solutions $\{\phi_\sigma^{(m)}\}$, corresponding to different configurations of single steps $\{\sigma_i\}$ and time-sequences $\{\tau_i\}$ for the tunneling events, form the basis in the well-known dilute instanton gas approximation.

At $T \ll T_s$, the multi-step solution (9) can be used as the extremal trajectory when calculating the partition function that appears in the expression (3) for the free energy. In this manner we get the zero temperature value of the persistent current as

$$I_{WC}(T=0) \sim (-1)^N \frac{es}{L} \left(\frac{T_s}{\alpha V_0} \right)^{1/\alpha} \sin(2\pi\Phi/\Phi_0). \quad (10)$$

This result for the persistent current of a Wigner crystal in the presence of a pinning potential barrier, clearly shows that the effect of the barrier is simply to suppress the zero temperature amplitude of the current. The net current depends on the elastic properties of the Wigner lattice that reflects the fact that for a strong pinning the charge transport in the ring is due to macroscopic quantum tunneling of the system through a deformed state of the crystal.

Except at very low temperatures, $T \ll T_s$, the only relevant saddle point trajectory is the single-step solution (8). By using it one gets for the normalized current

$$\frac{I_{WC}(T)}{I_{WC}(0)} = \frac{T}{T_s} \exp \left(\frac{1}{\alpha} f(T/T_s) \right), \quad T \ll \alpha V_0$$

$$f(x) = \frac{\pi}{2} x - \ln \left(\frac{\sinh(\pi x)}{\pi x} \right). \quad (11)$$

This result implies a non-monotonic temperature dependence of the persistent current; for a stiff crystal (small α) the current has an exponentially sharp maximum at $T \sim 0.5T_s$, with a width of the order of $\sqrt{T_0 T_s}$. The physical reason for this non-trivial temperature dependence – shown in Fig. 1 for different values of α – can be explained as follows: It is easy to see from (8) that as the temperature is increased the picture of the elastic deformation propagating as a “sharp” instanton changes (at $T \sim T_s$) into a picture of a homogeneous sliding of the crystal as a whole. This temperature-induced “softening” of the instanton reduces the contribution to the action from the elastic deformation of the crystal. Hence, the persistent current should increase with temperature. On the other hand this effect competes with a thermal smearing of the phase coherence which – as we showed for the unpinned crystal – tends to reduce the current. The sharp peak in the temperature dependence of the persistent current carried by a strongly pinned Wigner crystal, is a result of this very competition.

At weak pinning a stiff Wigner crystal-ring tunnels through the barrier as a whole and the persistent current does not depend on the elastic properties of the chain. However, if the barrier is high enough $\alpha T_s \lesssim V_p \ll T_s/\alpha$ (moderately weak pinning) the zero temperature current is

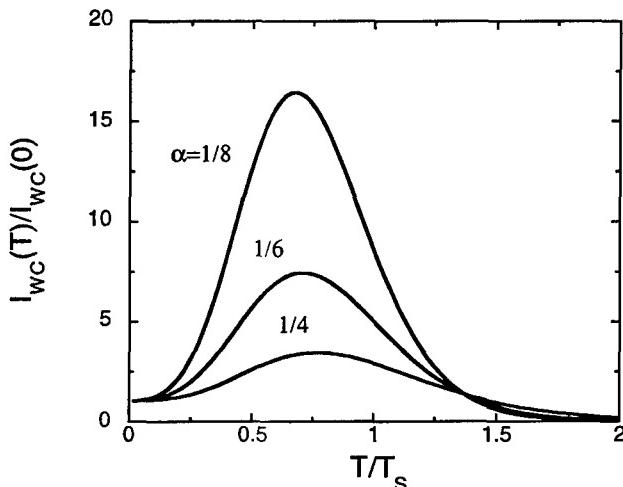


Fig. 1. Temperature dependence of the normalized persistent current in a strongly pinned Wigner crystal of different stiffness (measured by $\alpha^{-1} = 2msa/\hbar$; $T_s = \hbar s/k_B L$, see text). The sharp peak for stiff crystals is a result of a competition between two effects of temperature: a reduced renormalized tunneling barrier and a loss of quantum coherence.

still greatly suppressed,

$$I_{wp}(T \rightarrow 0) \sim (-1)^N I_0 \left(\frac{V_0}{T_0} \right)^{1/4} \times \exp \left(-\sqrt{2\pi} \frac{V_0}{T_0} \right) \sin \left(2\pi \frac{\Phi}{\Phi_0} \right). \quad (12)$$

This is in contrast to the case of non-interacting electrons where even a large potential barrier (of the order of the Fermi energy) only leads to a power-law suppression of the persistent current [17].

As temperature always makes tunneling easier, one may expect an anomalous temperature behaviour of the persistent current even for a weakly pinned Wigner crystal-ring. This is indeed the case, but unlike in the regime of strong pinning, the maximum of the persistent current (which is now attained at a potential-dependent temperature $T^* \sim \sqrt{V_0 T_0}$) is weakly pronounced. Therefore the only distinctive feature of the temperature dependence of a moderately pinned ($V_0 > T_0$) Wigner crystal-ring, – compared to that of a ring with noninteracting electrons – is the shift of the crossover temperature to higher values, $T_0 \rightarrow T^* > T_0$. For very weak pinning, $V_p \ll T_0$, the response of a Wigner crystal-ring to a magnetic flux is the same as for free electrons.

Formula (11) is valid in the strong pinning limit, when temperature is much smaller than αV_0 . At high temperatures, $T \gtrsim \alpha V_0$, the pinning potential can be treated as a perturbation when calculating the depinning of the Wigner crystal. In this case we find unimportant corrections to the persistent current in an ideal Wigner crystal (the details of this calculation will be published elsewhere [16]).

So far we have considered spinless electrons. Now we argue that for a stiff Wigner crystal-ring the results obtained within our model remain valid for electrons carrying spin as well. Note that spin degrees of freedom can affect the Aharonov-Bohm oscillations only via a plasmon-magnon interaction, which is strongly suppressed for a rigid crystal. In this case the electrons are well separated from each other and the overlap of the wavefunctions of neighbouring electrons is small. As was pointed out above, plasmon excita-

tions do not change the persistent current in a perfect Wigner crystal-ring at low temperatures, $T \ll E_F$. Therefore the effect of magnons via interactions with plasmons must also be small. In principle tunneling can create spin excitations. However, in our case the pinning potential is assumed to be smooth on the scale of the lattice spacing in the Wigner crystal, a , and therefore the overlap of wavefunctions of neighboring electrons does not change appreciably during the tunneling process. Hence, although the first tunneling step in the two-step model used in our calculation amounts to creating an excess charge e over a length l_0 , this process is not accompanied by the creation of any excess spin density; the subsequent relaxation involves only plasmons. In other words for a stiff Wigner crystal-ring the spin degrees of freedom completely decouple from the charge excitations and therefore do not contribute to the persistent current.

By measuring the dependence of the persistent current on the barrier height at zero temperature (10) and its temperature dependence (11), one has an opportunity to determine independently the stiffness parameter, $\alpha = h/2msa$, and the sound velocity, s , in this system of strongly correlated electrons. This gives us strong reasons to propose an experiment using a gate-controlled barrier in a mesoscopic semiconductor ring in order to study Wigner crystallisation and to measure the parameters of the crystal.

In conclusion we have shown that in an ideal ring with no impurity scattering, the persistent current carried by interacting electrons – so strongly correlated that they form a Wigner crystal – is indistinguishable from the current carried by a non-interacting Fermi gas. By incorporating a potential barrier, $V_0 > \hbar v_F/L$, in the ring structure, a qualitative change of the magnitude and temperature dependence of the persistent current appears. With an adjustable barrier, these differences can be used for detecting and investigating the properties of the Wigner crystal.

Acknowledgements

We gratefully acknowledge discussions with L. Glazman, A. Nersesian, A. Sjölander, and A. Zagorskin. This work was supported by the Swedish Royal Academy of Sciences, the Swedish Natural Science Research Council, the Swedish National Board for Industrial and Technical Development, by the NSF through grant DMR-9113911, and by grant UK000 from International Science Foundation. One of us (I.K.) acknowledges the hospitality of the Department of Applied Physics, CTH/GU.

References

1. Mailly, D., Chapelier, C. and Benoit, A., Phys. Rev. Lett. **70**, 2020 (1993).
2. Liu, J., Gao, W. X., Ismail, K., Lee, K. Y., Hong, J. M. and Washburn, S., Phys. Rev. B (unpublished).
3. Typically only few transverse electronic modes exist in such a system similarly to what one has in quantum point contacts and quantum wires. Therefore it is reasonable to call them quantum rings.
4. Günter, L. and Imry, Y., Solid State Commun., **7**, 1391 (1969).
5. Kulik, I. O., JETP Lett. **11**, 275 (1970).
6. Ambegaokar, V. and Eckern, U., Phys. Rev. Lett. **65**, 381 (1990).
7. Müller-Groeling, A., Weidenmüller, H. A., and Lewenkopf, C. H., Europhys. Lett. **22**, 193 (1993).
8. Larkin, A. I. and Lee, P. A., Phys. Rev. B**17**, 1596 (1978).
9. Glazman, L. I., Ruzin, I. M. and Shklovskii, B. I., Phys. Rev. B**45**, 8454 (1992).
10. Kane, C. L. and Fisher, M. P. A., Phys. Rev. B**46**, 15233 (1992).
11. Krive, I. V. and Rozhavsky, A. S., Int. J. Mod. Phys. B**6**, 1255 (1992).

12. Byers, N. and Yang, C. N., Phys. Rev. Lett., **7**, 46 (1961).
13. Loss, D., Phys. Rev. Lett. **69**, 343 (1992).
14. Leggett, A. J., in: "Granular Electronics" (Edited by D. K. Ferry, J. R. Barker and C. Jacoboni), NATO ASI Ser. B, vol. 251 (Plenum, New York, 1991), p. 297.
15. Bogachev, E. N., Krive, I. V., Kulik, I. O. and Rozhavsky, A. S., Phys. Rev. **B42**, 7614 (1990); Sov. Phys. JETP, **70**, 336 (1990).
16. Krive, I. V., Shekhter, R. I., Girvin, S. M. and Jonson, M. (unpublished).
17. Cheung, H.-F., Gefen, Y., Riedel, E. K. and Shih, W.-H., Phys. Rev. **B37**, 6050 (1988).
18. Grincwajg, A., Jonson, M., Krive, I. V., and Shekhter, R. I. (unpublished).
19. This instanton solution is analogous to the "colorons" in quantum chromodynamics. Harrington, B. and Shepard, H., Phys. Rev. **D17**, 2122 (1978); Gross, D. J., Pisarski, R. D. and Yaffe, L. G., Rev. Mod. Phys. **53**, 43 (1981).

Unified Capacitance Modelling of MOSFETs

O. G. Johannessen, T. A. Fjeldly* and T. Ytterdal

Department of Physical Electronics, Norwegian Institute of Technology, University of Trondheim, N-7034 Trondheim, Norway

Received April 28, 1994; accepted June 14 1994

Abstract

A unified physics based capacitance model for MOSFETs suitable for implementation in circuit simulators is presented. This model is based on the charge conserving, so-called Meyer-like approach proposed by Turchetti *et al.*, and utilizes a unified charge control model to assure a continuous description of the MOSFET capacitances both above and below threshold. The capacitances associated with the model are comparable to those of the well-known BSIM model in the above-threshold regime, but it is more precise in the description of near-threshold and subthreshold behaviour. Moreover, the discontinuities at the transitions between the various regimes of operation are removed. The present modelling scheme was implemented in our circuit simulator AIM-Spice, and simulations of the dynamic behaviour of various demanding benchmark circuits clearly reveal its superiority over simulations using the simple Meyer model.

1. Introduction

Accurate modelling of MOSFET capacitance-voltage (*C-V*) characteristics is needed for precise simulation of transient and small-signal behaviour of MOSFET/CMOS circuits. Previously, we introduced a modification of the simple *C-V* model by Meyer [1] in order to account for subthreshold conditions, using the concept of a unified channel capacitance [2]. However, a serious drawback of the Meyer model is its use of reciprocal capacitances between the terminals which causes a violation of charge conservation. This problem was addressed by the introduction of the charge conserving, non-reciprocal capacitance model by Ward and Dutton [3]. (In fact, the Meyer capacitances are a subset of the Ward-Dutton capacitances in the quasi-static approximation.) However, this model is relatively complex for application in standard transient circuit analyses.

Combining the Unified Charge Control Model (UCCM) [4] of FETs with a charge based model description, utilizing a precise formulation of the bulk charge in terms of body plot parameters, we were able to derive a more accurate, charge conserving *C-V* model which simultaneously accounts for non-uniform doping profiles, subthreshold behaviour, and short channel effects [2, 5]. But this model is also quite cumbersome to implement in SPICE since it does not yield analytical expressions for the charge distribution in the channel.

Here, we describe yet another unified *C-V* model for MOSFETs, suitable for implementation in SPICE. This model utilizes an approximate, analytical solution of UCCM in combination with the charge conserving, Meyer-like approach proposed by Turchetti *et al.* [6] – all within the quasi-static approximation. This approach relies on the use of the Meyer capacitances within a so-called companion model which, through the use of a charge conserving iter-

ation procedure during the SPICE simulation, yields the correct charges and terminal currents in the convergence limit. The capacitances are derived on the basis of analytical expressions for the inversion and depletion charges, similar to those obtained by Iñiguez and Moreno [7, 8]. In addition, in order to assure charge conservation, a so-called charge partitioning scheme must be enforced whereby the distributed inversion charge is partitioned between the source and the drain terminals. The complete model was implemented in our circuit simulator AIM-Spice [2].

In this paper, we evaluate the present model by (i) comparing the present unified capacitances with their corresponding Ward-Dutton (W-D) counterparts as implemented in the well-known BSIM model, and (ii) by simulating various demanding benchmark test circuits such as charge pumps, switched capacitor circuits and dynamic memory cells. These tests clearly reveal shortcomings of the original Meyer model and of the BSIM W-D model.

2. Model specifics

The following steps are used for developing the MOSFET capacitance model:

- Express the inversion charge sheet density $-qn_i$ in the channel in terms of a charge control model.
- Use the basic drift-diffusion equations to formulate the drain current I_d in terms of n_i and integrate this equation to find I_d as a function of the terminal voltages.
- Integrate $-qn_i$ over the gate area to obtain the total inversion charge Q_I .
- Partition Q_I between a “source charge” Q_S and a “drain charge” Q_D .
- Express the depletion sheet charge density $-qn_b$ in terms of the potential drop ψ_{sB} across the semiconductor (between the oxide interface and the substrate interior) using the depletion approximation; write the gate-substrate voltage as $V_{GB} = \psi_{sB} + \psi_{ox}$, where ψ_{ox} is the potential drop across the oxide; write the gate sheet charge density as $qn_g = qn_i + qn_b$; use Gauss’ law to write $\psi_{ox} = qn_g/c_{ox}$, where c_{ox} is the oxide capacitance per unit area of the gate. Eliminate ψ_{sB} and ψ_{ox} between these relationships to derive an expression for n_b in terms of V_{GB} and n_i .
- Integrate $-qn_b$ over the gate area to obtain the total depletion charge Q_B .
- Calculate the capacitances as appropriate derivatives of charges with respect to terminal voltages. For example, the following are the Meyer capacitances used here:

$$C_{GS} = \frac{\partial Q_G}{\partial V_{GS}}, \quad C_{GD} = \frac{\partial Q_G}{\partial V_{GD}}, \quad C_{GB} = \frac{\partial Q_G}{\partial V_{GB}} \quad (1)$$

where $Q_G = -(Q_I + Q_B)$.

* Corresponding author.

In the Unified Charge Control Model, n_i is related to the gate voltage V_{GS} and the local quasi-Fermi potential V_F by the expression [2, 4]

$$V_{GS} - V_T - \alpha V_F = \eta V_{th} \ln(n_i/n_o) + a(n_i - n_o) \quad (2)$$

where V_T is the threshold voltage, α is a constant close to unity that accounts for a shift in threshold voltage with position in the channel, η is the subthreshold ideality factor, $V_{th} = k_B T/q$ is the thermal voltage, n_o is the inversion charge density at threshold, and $a \approx q/c_{ox}$.

Following a straightforward derivation for a long-channel MOSFET, we arrive at the following results for the inversion charge and the depletion charge [7, 8], respectively:

$$Q_I = -qWL \frac{\eta V_{th}(n_{id} + n_{is}) + \frac{2a}{3}(n_{id}^2 + n_{id}n_{is} + n_{is}^2)}{2\eta V_{th} + a(n_{id} + n_{is})} \quad (3)$$

$$Q_B \approx -qWLn_{bs} - \frac{qWLn_{is} + Q_I}{1 + 2an_{bs}/\gamma^2} \quad (4)$$

where $-qn_{id}$ and $-qn_{is}$ are the inversion charge densities evaluated from eq. (2) at source and the drain, respectively, and $-qn_{bs}$ is the depletion charge density at source given by

$$n_{bs}(n_{is}) = \frac{\gamma}{2} \sqrt{\left(\frac{\gamma}{a}\right)^2 + \frac{4}{a^2}(V_{GB} - V_{FB})} - \frac{4}{a}n_{is} - \frac{\gamma^2}{2a} \quad (5)$$

Here, V_{GB} is the gate-substrate voltage, V_{FB} is the flat-band voltage, and $\gamma = \sqrt{2q\epsilon_s N_d/c_{ox}}$ is the body effect constant where N_d is the doping density and ϵ_s is the dielectric permeability in the substrate.

In order to obtain analytical expressions for the charges Q_I and Q_B and the Meyer capacitances (or the W-D capacitances), we need a solution of eq. (2) with respect to n_i . Since this equation cannot be solved analytically, we propose the following approximate solution [2, 9]

$$n_i \approx 2n_o \ln \left[1 + \frac{1}{2} \exp \left(\frac{V_{GS} - V_T - \alpha V_F}{\eta V_{th}} \right) \right] \quad (6)$$

This approximation is in good agreement with UCCM everywhere and has the correct asymptotic behaviour both above and below threshold. Hence, analytical expressions for n_{is} and n_{id} are obtained by using $V_F = 0$ and $V_F = V_{DS}$, respectively, in eq. (6).

3. Model evaluation

Although the Meyer capacitances are used in the SPICE Newton-Raphson algorithm, they do not enter in the final results of the computational iteration procedure. Instead, the results are determined by the charge model used, such as for examples eqs (2)–(6). However, the capacitances derived from these charge models by means of eq. (1) give a better physical understanding of the quality of the model.

The three Meyer capacitances are, in the quasi-static approximation, identical to three of the nine W-D capacitances. Figure 1 shows a comparison between the present unified Mayer capacitances and the corresponding W-D capacitances as modelled and implemented in BSIM. We notice that the overall agreement between these two models is quite good, except that the W-D BSIM model describes poorly the near-threshold and subthreshold regimes. While

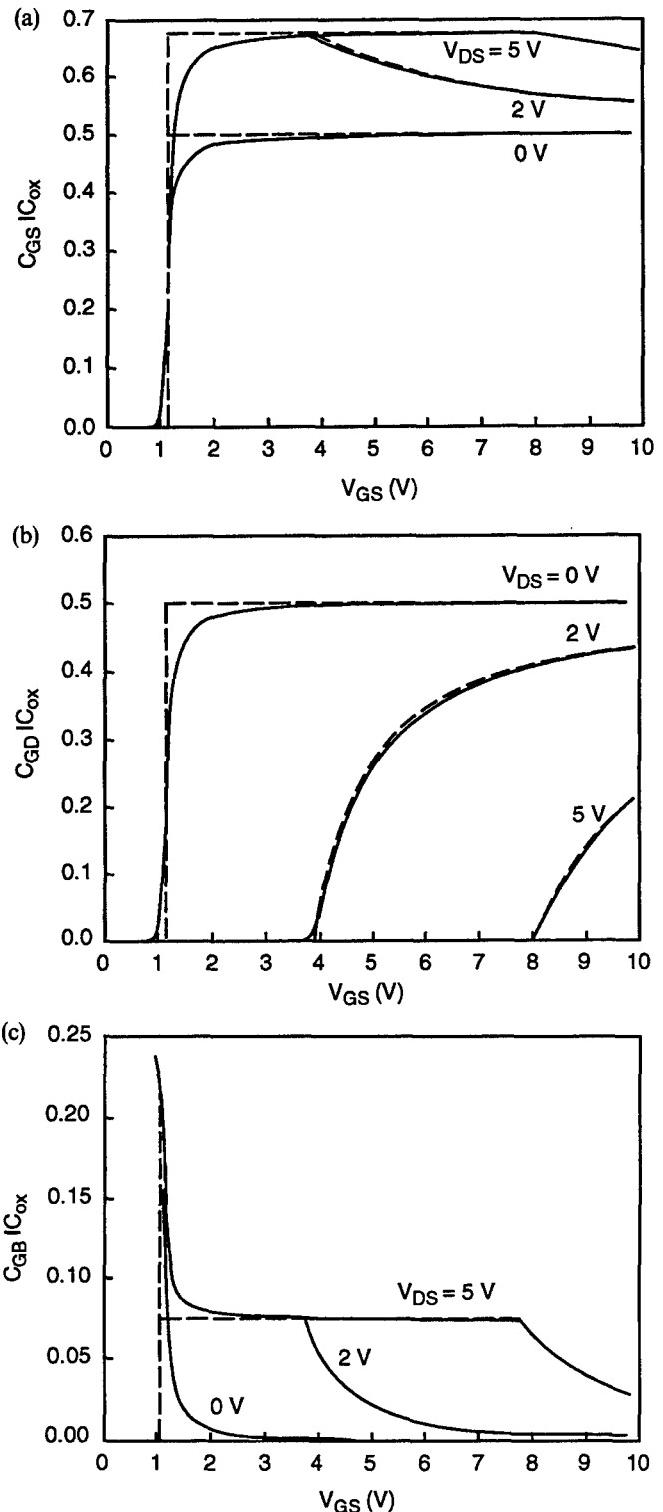


Fig. 1. Comparisons of the normalized capacitances C_{GS}/C_{ox} (a), C_{GD}/C_{ox} (b), C_{GB}/C_{ox} (c), calculated from the present unified model (solid lines) with the corresponding Ward-Dutton capacitances used in BSIM (dashed lines).

the present capacitances are continuous and smooth everywhere, in agreement with experiments [2, 5], the non-unified BISM W-D capacitances are characterized by discontinuities at threshold and discontinuous first derivatives with respect to V_{GS} at the saturation voltage. Such discontinuities invariably lead to a reduction in convergence and computational speed in SPICE simulations [10].

The importance of using charge conserving MOSFET capacitance modelling in SPICE simulations was demonstrated for different benchmark circuits. These circuits, which include charge pumps, switched capacitor filters and

Dynamic Random Access Memory (DRAM) cells, were selected because of their sensitivity to charge non-conservation. In all cases, the Meyer-like approach gave results which correspond very well with analytical estimates, while the simple Meyer model often resulted in significant deviations. For example, in Fig. 2 we show AIM-Spice simulation results for a simple charge pump excited by a periodic pulse train at the input. We notice that the deviation in the output using Meyer's model is large and somewhat erratic as a result of charge non-conservation.

Another example is shown in Fig. 3, where we simulate the operation of a switched capacitor low-pass network. Again, the results for the Meyer model deviates significantly and erratically from the expected behaviour, while the latter was very well reproduced by the Meyer-like approach.

Our investigation of a three-transistor DRAM cell revealed that the use of Meyer's model resulted in failure to identify a potentially serious down-shift in the DATA output signal (from DATA = 1) during *read* operations. This effect is caused by so-called clock feed-through and charge injection into the data holding capacitor, and a very precise capacitance modeling incorporating both charge conservation and a proper description of subthreshold behaviour is required in order to characterize the effect satisfactorily and to establish a correct feedback to the design.

3. Summary

We have presented a unified and charge conserving capacitance modelling scheme for MOSFETs, for use in circuit simulators. The method is based on the Unified Charge Control Model, in combination with the so-called Meyer-

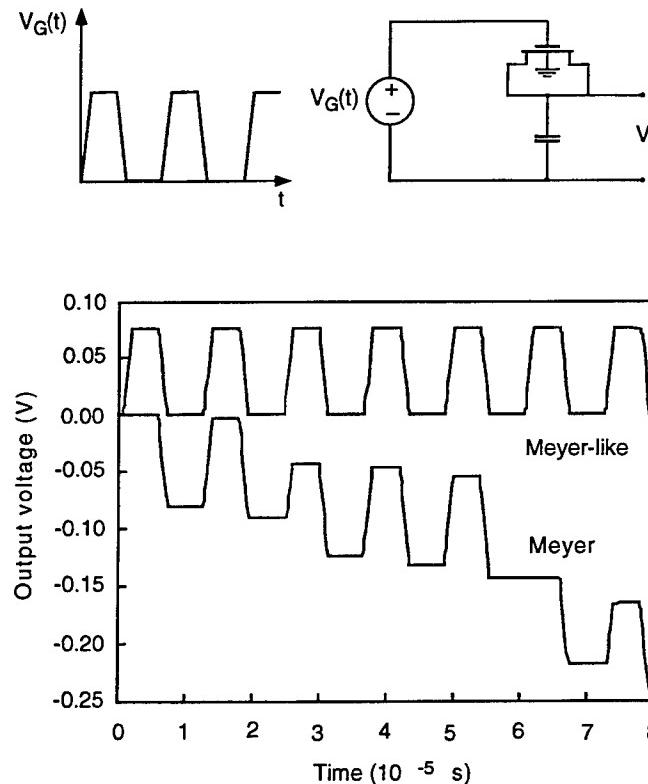


Fig. 2. Comparison of simulated output from a simple charge pump (shown on top) using the present modelling scheme (Meyer-like) and the simple Meyer model (Meyer).

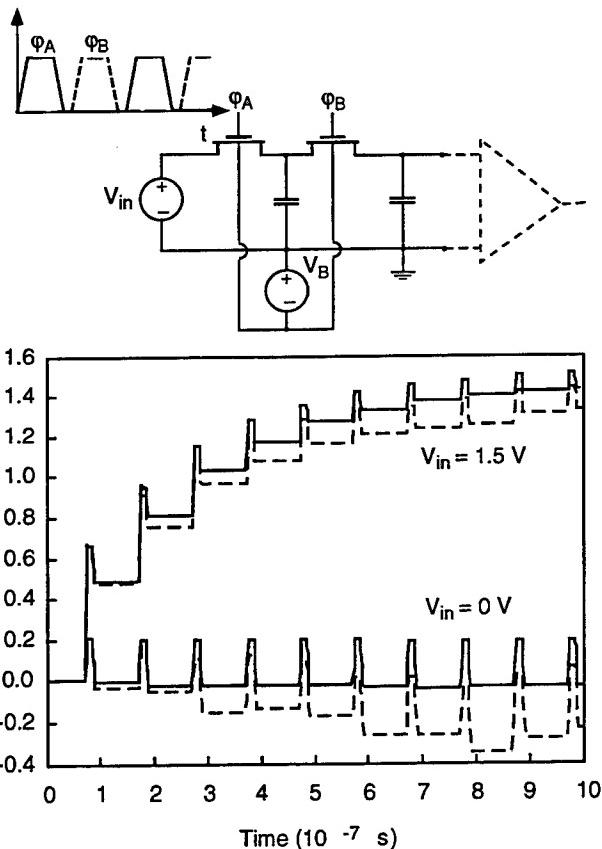


Fig. 3. Comparison of simulated outputs from a switched capacitor low-pass filter (shown on top) using the present Meyer-like modelling scheme (solid curves) and the simple Meyer model (dashed curves).

like SPICE implementation. The capacitances associated with this model compare well with the corresponding Ward-Dutton capacitances implemented in the well-known BSIM model, but are more precise in the near-threshold and in the subthreshold regimes and do not have discontinuities at the transitions between different regimes of operation.

The present modelling scheme was implemented in our circuit simulator AIM-Spice and was used for test simulations of several benchmark circuits. The test results show that the present modelling approach reproduces the expected results very well, while simulations based on the simple Meyer model show significant deviations, owing to the lack of charge conservation.

References

1. Meyer, J. E., *RCA Rev.* **32**, 42 (1971).
2. Lee, K., Shur, M., Fjeldly, T. A. and Ytterdal, T., "Semiconductor Device Modeling for VLSI" (Prentice Hall, New Jersey 1993).
3. Ward, D. E. and Dutton, R. W., *IEEE J. Solid-State Circuits* **SC-13**, 703 (1978).
4. Byun, Y., Lee, K. and Shur, M., *IEEE Electron Device Letters EDL-11*, 50 (1990).
5. Rho, K-M., Lee, K., Shur, M. and Fjeldly, T. A., *IEEE Trans. Electron Devices ED-40*, 131 (1993).
6. Turchetti, C., Prioretti, P., Masetti, G., Profumo, E. and Vanzi, M., *IEEE Trans. on Computed-Aided Design CAD-5*, 499 (1986).
7. Iñiguez, B. and Moreno, E. G., *Electronic Letters* **29**, 1036 (1993).
8. Iñiguez, B. and Moreno, E. G., "Proceedings of ISCAS'94" (London 1994), to be published.
9. Fjeldly, T. A., Moon, B-J. and Shur, M., *IEEE Trans. Electron Devices TED-38*, 1796 (1991).
10. Ytterdal, T., Fjeldly, T. A., Shur, M. and K. Lee, "Proceedings of ISDRS'93" (Charlottesville, Virginia 1993), p. 345.

GaAs/AlGaAs Quantum Well Infrared Photodetector Arrays for Thermal Imaging Applications

L. Lundqvist, J. Y. Andersson and J. Borglind

Industrial Microelectronics Center (IMC), P.O. Box 1084, S-16421 Kista, Sweden

Received May 3, 1994; accepted June 14, 1994

Abstract

The performance of GaAs/AlGaAs multiple quantum well infrared detectors is studied theoretically and experimentally, with special emphasis on 8–12 μm thermal imaging applications. The dependence of detector performance on various factors like light coupling configurations (one and two dimensional reflection gratings or 45° polished edge), detector temperature, response wavelength and quantum well doping density is dealt with. An absorption quantum efficiency of 87% is demonstrated using a crossed grating and a waveguide (CGW). It is also found that an optimised 34 $\mu\text{m} \times 34 \mu\text{m}$ detector (a detector size suitable for large staring arrays, i.e. 256 × 256 or larger) with 9.0 μm cut-off wavelength, f# = 2 optics and 70% optical transmission reaches background limited operation at 74 K detector temperature.

The potential of making highly uniform staring arrays utilising the mature GaAs material and processing technology is demonstrated by uniformity measurements of detector dark current. The experiments show that a metalorganic vapour phase epitaxy (MOVPE) grown structure can have a dark current standard deviation to mean value ratio over a 10 mm long linear detector array of less than 2%.

The staring array performance in terms of noise equivalent temperature difference (temporal NETD) is calculated to NETD < 20 mK at 77 K detector temperature and NETD < 10 mK at 70 K detector temperature.

1. Introduction

Since the first demonstration of a GaAs/AlGaAs quantum well infrared photo detector (QWIP) [1] the physics and performance of QWIPs have been thoroughly studied [2]. Normally, semiconductor materials with small bandgaps E_g ($E_g \sim 0.1\text{--}0.2 \text{ eV}$) such as HgCdTe, enabling optical detectors active in the 8–12 μm long wavelength infrared atmosphere transfer window (the LWIR window), are more difficult to grow, process and fabricate into uniform devices than are larger band-gap ($E_g > 1 \text{ eV}$) materials [3]. One of the main driving forces for QWIP technology development is the ability to create semiconductor devices with a small bandgap optical behaviour combined with a large bandgap mechanical and chemical stability. Therefore, GaAs/AlGaAs QWIP technology is a viable candidate for large, high performance, low cost LWIR (8–12 μm) focal plane arrays (FPAs) [4–6].

QWIPs operate on account of intersubband transitions in doped quantum wells (QWs) which implies photoexcitation of electrons (holes) from a bound ground state E_1 (H_1) to a quasi-bound or extended excited state E_2 (H_2), see Fig. 1. Excited charge carriers at E_2 (H_2) are freely mobile perpendicularly to the QW planes, thus enabling photoconductive action. As a result of low effective mass, QWIPs with n-doped QWs offer the highest detectivity D^* . The drawback with n-doped QWIPs is that the quantum mechanical selection rules forbid absorption of radiation with incidence normal to the QW-layer plane and much work has been

done to find optical geometries that overcome this fact and enhance the absorption quantum efficiency (absorptance) η of the detectors [7–13].

One of the most efficient optical geometries presented in the crossed (doubly periodic) grating with a waveguide (CGW) [10]. The crossed grating is etched into the top of the detector mesa and the waveguide is defined by the grating, the active QW layer and a cladding layer, from top to bottom, see Fig. 2. In this work we compare the CGW geometry with the standard 45° polished edge (EDGE) geometry. We also investigate the influence of QW sheet

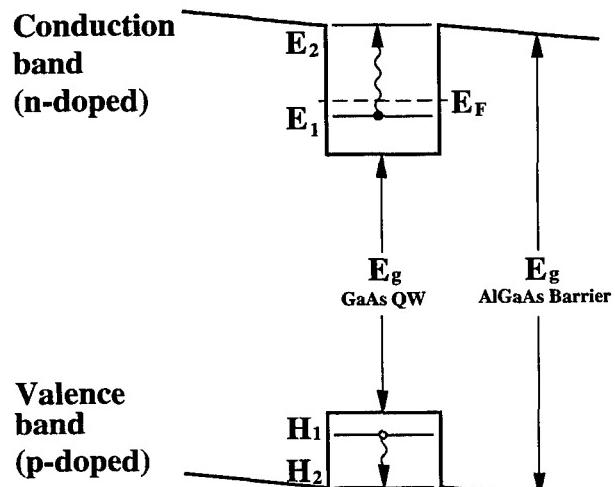


Fig. 1. Energy band structure of a GaAs/AlGaAs quantum well (QW). Intersubband absorption between E_1 and E_2 (electrons) or H_1 and H_2 (holes) is schematically shown. The photo generated charge carriers at E_2 or H_2 are under the influence of an applied electric field giving rise to a photo current. The Fermi energy level E_F is indicated (for the case of an n-doped QW).

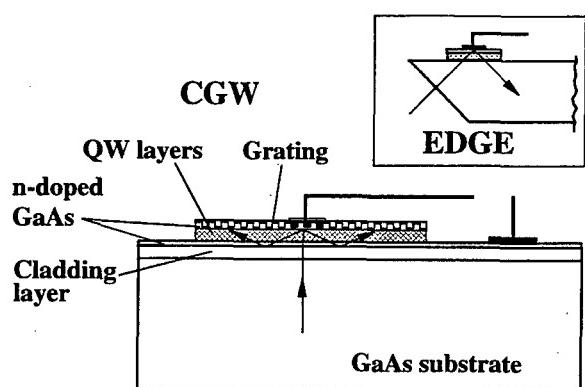


Fig. 2. Cross section of a detector with an etched grating and waveguide (CGW) as well as a standard EDGE detector.

doping concentration, optical response wavelength and detector temperature on QWIP performance. Finally we use the experimental single QWIP results to calculate the performance of detector arrays in terms of noise equivalent temperature difference (NETD) and photo generated current to thermally generated dark current ratio (N_{ph}/N_{th} ratio).

2. Experimental

Four different structures A, B, C and D are grown onto GaAs semi-insulating substrates using a low pressure metal-organic vapour phase epitaxy (MOVPE) reactor featuring wafer rotation. All structures contain a 50 period (GaAs/Al_xGa_{1-x}As) multi-QW stack sandwiched between a 1.0 μm bottom and a 1.3 μm top GaAs contact layer ($n = 7 \cdot 10^{17} \text{ cm}^{-3}$) and an additional 3 μm Al_{0.80}Ga_{0.20}As cladding layer located between the substrate and bottom contact. The measured structure parameters, QW width a_{QW} , barrier width a_{bar} , barrier Al-content x_{Al} , QW doping sheet concentration n_{QW} are summarised in Table I. Structure A, B and C have different QW doping sheet concentration n_{QW} whereas structure D has a lower barrier Al-content x_{Al} .

The detectors are fabricated as follows: gratings (linear, crossed with square symmetry, or no grating) are etched by reactive ion etching (RIE) into the GaAs top contact layer. The grating depth $h = 0.80 \mu\text{m}$ throughout, while the other grating parameters are fabricated with different sizes ranging from grating constant $D = 2.3$ to $3.0 \mu\text{m}$, and crossed grating box shape cavity width $d \approx 0.6D$ and linear grating cavity width $d \approx 0.5D$. For the definition of the grating parameters see Fig. 3 and [12]. Square shaped mesas of various sizes (from 25 to 500 μm side length) are defined by wet etching down to the bottom contact layer. AuGe/Ni ohmic contacts are deposited onto the top

(covering ~10% of the grating area) and bottom contact layer and finally a reflective Au layer is sputtered covering the whole grating area. For characterisation of the detectors, a grating monochromator, a 1000 K glowbar source and a pyroelectric reference detector are used. A 1000 K black-body source is exploited for determining black-body responsivities and detectivities as well as for calibration purposes.

3. Single detector performance

The detector current I_d in QWIPs consists of two parts, the photo generated current I_{ph} and the undesired thermally generated dark current I_{th} . I_{ph} can be described by

$$I_{ph} = \int_{\lambda_1}^{\lambda_2} R_I P_{opt} d\lambda$$

where R_I is the spectral current responsivity and P_{opt} the spectral optical power illuminated on the detector. In this work we will set λ_1 and λ_2 to $8 \mu\text{m}$ and $12 \mu\text{m}$ respectively, since we are interested in the radiation generated in the LWIR window. The thermally generated dark current I_{th} has been investigated and modelled by several authors [2] and here we will use the model presented in [14]. The most important parameters concerning the relationship between I_{ph} and I_{th} are n_{QW} , detector temperature T , cut off wavelength λ_c (or cut off energy E_c) and the method of optical coupling. This is described by,

$$I_{ph} \propto R_I \propto \eta = 1 - e^{-c_1 \cdot n_{QW}}$$

where c_1 is a constant invoking the method of optical coupling,

$$I_{th} \propto T \cdot \exp - \left(\frac{E_c - E_F}{kT} \right),$$

where $E_c = hc/\lambda_c$ and the Fermi level energy

$$E_F \propto T \cdot \ln [e^{c_2 \cdot n_{QW}} - 1]$$

where c_2 is a constant. Response measurements of different types of detectors are carried out and the results are presented in Table II. The absorption quantum efficiency (absorptance) η are calculated using two independent methods.

The first method (applied to structure A and C) is based on the measurement of absorptance with a Fourier transform infrared spectrometer using a 45° polished edge geometry (EDGE), and subsequently the current response R_I of a detector with the same optical coupling geometry is measured. From

$$R_I = \eta g p_e \cdot \frac{q\lambda}{hc},$$

where R_I is the current responsivity, h the Planck constant, c the vacuum speed of light, q the electron charge, g the photoconductive gain and p_e the probability of an excited electron to escape from the QW, we can calculate $g p_e$. Knowing $g p_e$ it is possible to derive the absorptance, henceforth called η_{FTIR} for other types of coupling geometries (i.e. grating detectors) simply by measuring R_I .

The second method (applied to structure A, B, C and D) is based on deriving g from the detector generation-recombination (g-r) noise current i_{g-r} . The noise current i_d

Table I. Structural parameters for the four types of multi QW IR detectors. The parameters are obtained from X-ray diffraction rocking curves and C-V etch profiling

Structure	QW width, a_{QW} (Å)	Barrier width, a_{bar} (Å)	Barrier Al- content, x_{bar} (%)	QW doping concentration, $n_{QW} \times 10^{11}$ (cm $^{-2}$)
A	49	339	29.8	1.6
B	48	337	29.8	2.5
C	50	339	30.2	4.1
D	52	348	29.0	1.9

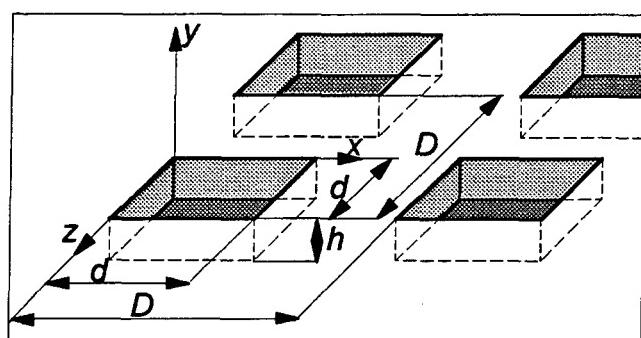


Fig. 3. Detail of the geometry of the crossed grating viewed from the GaAs side. D is the grating constant and d the cavity width.

Table II. Single detector data for the structures A, B, C and D. The detector type (light coupling method) are either of the type 45° polished edge (EDGE), or crossed grating and waveguide (CGW (D)) where (D) is the grating constant. The measurements have been conducted at 77 K on 150 × 150 μm square detector mesas with a detector bias of -4.0 V. The g-r noise currents are measured at 77 K background temperature. The two types of photoconductive gain (g_{FTIR} , g_{g-r}) and absorptance (η_{FTIR} , η_{g-r}) refer to the two different methods of obtaining absorptance described in the text, respectively

Structure	Detector type	Peak wavelength, λ_{peak} , (μm)	Dark current, I_{th} (nA)	g_{g-r} noise current, i_{g-r} (fA/ $\sqrt{\text{Hz}}$)	Peak current responsivity, R_I (mA/W)	Peak detectivity, $D^*(\times 10^{11})$ (cm $\sqrt{\text{Hz W}^{-1}}$)	Photoconductive gain, g_{FTIR}/g_{g-r} (-)	Absorptance, η_{FTIR}/η_{g-r} (%)
A	EDGE (-)	8.3	—	—	84	—	0.16/—	7.7/—
A	CGW (2.6)	8.3	10	31	570	2.8	0.16/0.15	52/57
B	CGW (2.6)	8.3	29	52	750	2.2	—/0.15	—/77
C	EDGE (-)	8.3	—	—	140	—	0.11/—	20/—
C	CGW (2.6)	8.3	57	64	640	1.5	0.11/0.11	91/87
D	CGW (2.8)	8.8	46	72	630	1.3	—/0.18	—/49

present in QWIPs is described by

$$i_d^2 = i_{g-r}^2 + i_J^2 + i_{1/f}^2$$

where the Johnson noise current $i_J = \sqrt{4kT/dR}$, k the Boltzman constant, T the detector temperature and dR the detector dynamic resistance. Properly fabricated, the detector 1/f noise current $i_{1/f}$ is negligible. At normal operating conditions i_{g-r} is the dominating noise current. By using

$$g = i_{g-r}^2/(4qI_d),$$

(where I_d is the DC detector current) the fact that $p_e \sim 1$ at high enough detector bias, and the measured R_I it is possible to obtain the absorptance, henceforth labelled η_{g-r} .

Table II shows that the two methods are in good agreement within ~5%. When comparing the CGW and the standard EDGE light coupling methods, the CGW absorptance is a factor of 7 higher (structure A). For structure C, this factor is as low as 4.5 indicating that the CGW absorptance is near saturation (i.e. close to unity). As a result of increased QW sheet doping concentration, the CGW absorptance η_{g-r} for structure A, B, and C has a rising trend from 57%, 77% to 87%. Detectors with linear gratings and

waveguide (LGW) have roughly a factor of 2 lower absorptance [12]. The spectral absorptance is shown in Fig. 4.

The most important figure of merit for single IR detectors is a normalised signal to noise ratio, the detectivity D^* , obtained according to,

$$D^* = \frac{R_I}{i_d} \cdot \sqrt{A_d \Delta f},$$

where A_d is the optical detector area and Δf the noise bandwidth. Values of D^* for the different structures and light coupling geometries are presented in Table II. Structure A (lowest doping, lowest absorptance!) has the largest $D^* = 2.8 \cdot 10^{11}$ Jones. It is evident that high absorptance alone is not the only factor to consider for QWIP optimisation. For two structures with similar QW doping, a shift in peak wavelength λ_{peak} from 8.3 μm (structure A) to 8.8 μm (structure D) leads to a decrease in D^* of ~2. This exemplifies the common knowledge that λ_{peak} should be as short as the application permits and the difference between λ_{peak} and λ_c small.

4. Calculation of the performance of detector arrays

It is a straightforward procedure to arrange QWIPs into a focal plane array of chosen size. Desirable array sizes can be 128 × 128, 256 × 256, 320 × 240 or larger. The usual way to address each detector pixel is to flip-chip mount (via In-bumps) the detector array to a silicon multiplexer readout circuit. It is common practice for focal plane arrays to use the total number of collected charge carriers N or the number of noise electrons n , during an integration period T_{int} , instead of detector current and noise current. This is especially convenient when electrons are physically collected in an integration capacitor, as for the case of direct injection readout circuits. The relation between N (or n) and either direct current or noise current I is: $N = I \cdot T_{\text{int}}/q$. The integration time T_{int} is assumed to be 17 ms below, corresponding to a 60 Hz frame rate.

For detector arrays the noise equivalent temperature difference NETD is a suitable measure of performance especially for thermal imaging applications. A calculation of NETD of detector arrays is performed below, taking measured detector data from single detectors with mesa size

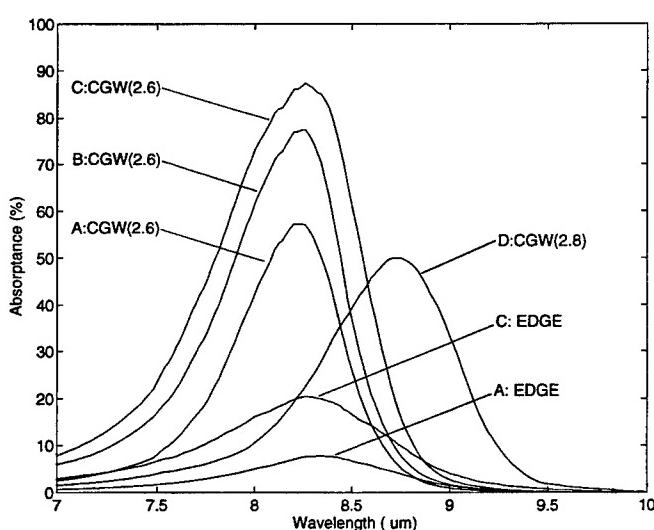


Fig. 4. Absorptance vs. wavelength for 150 × 150 μm detectors. The detectors with a crossed grating and waveguide (CGW) have a grating constant $D = 2.6 \mu\text{m}$ (structure A, B, C) or $D = 2.8 \mu\text{m}$ (structure D). The absorptance for the standard 45° polished edge (EDGE) optical geometry is included for structure A and C.

$34\text{ }\mu\text{m} \times 34\text{ }\mu\text{m}$ which is a suitable size for 256×256 and 320×240 arrays.

The total noise, n_{tot} associated with arrays consists of temporal noise n_{temp} and fixed pattern noise n_{fix} according to,

$$n_{\text{tot}}^2 = [n_{\text{temp}}^2] + n_{\text{fix}}^2 = [n_d^2 + n_{ro}^2] + n_{\text{fix}}^2,$$

where n_d is the detector noise and n_{ro} the readout circuit noise.

The detector noise is described by,

$$n_d^2 = 2gN_d + n_J^2 + n_{1/J}^2$$

where N_d is the total number of charge carriers (electrons) which equals the number of optically generated electrons N_{ph} and the number of thermally generated electrons N_{th} . For a well designed readout circuit, n_{ro} should be smaller than n_d . In the calculations below it is assumed that n_{ro} can be neglected.

The fixed pattern noise n_{fix} , arises as a result of non-uniformity of detector dark current and responsivity of the detector array. For the case of QWIP, n_{fix} is the remaining number of noise electrons after two point compensation of the detector current (e.g. offset and gain correction), and can be written [15]:

$$n_{\text{fix}}^2 = \beta_{\text{gn}}^2 \cdot (N_{\text{op}} \sigma_{\text{op}})^2 + \beta_{\text{of}}^2 \cdot [(N_{\text{th}} \sigma_{\text{th}})^2 + (N_{\text{op}} \sigma_{\text{op}})^2]$$

where σ_{op} and σ_{th} are the ratios of standard deviation to mean of photon current and dark current, respectively. The factors β_{of} and β_{gn} represent offset and gain corrections determined by sensor calibration. A proper calculation of n_{fix} requires knowledge of the uniformity of calibration sources in order to estimate β_{of} and β_{gn} . In the calculations below we will omit this contribution thus invoking only the temporal noise contribution. Nevertheless, n_{fix} is in most cases a dominating noise component and it should be noted that n_{fix} is linearly dependent on σ_{op} and σ_{th} making evident the importance of highly uniform arrays. We have flip chip mounted 128×128 arrays to commercial silicon multiplexers and used them in thermal imaging cameras. They show that $\sigma_{\text{op}} \cdot \sqrt{\beta_{\text{gn}}^2 + \beta_{\text{of}}^2} \approx 5 \cdot 10^{-4}$ for a 20 K calibration interval. We have also measured the 77 K σ_{th} on linear arrays, 1×81 with $110\text{ }\mu\text{m}$ mesas ($120\text{ }\mu\text{m}$ pitch) and 1×128 with $45\text{ }\mu\text{m}$ mesas ($50\text{ }\mu\text{m}$ pitch), which resulted in $\sigma_{\text{th}} = 1.8\%$ and $\sigma_{\text{th}} = 2.0\%$ as best results. σ_{op} is expected to be slightly higher [9]. These values represent state of the art uniformity for LWIR detector arrays and clearly demonstrates the advantages with QWIP technology.

The temporal noise equivalent temperature difference (NETD_{temp}) of the detector array is obtained from:

$$\text{NETD}_{\text{temp}} = \frac{1}{C_T} \cdot (n_{\text{temp}}/N_{\text{ph}})$$

where C_T is the differential temperature contrast according to,

$$C_T = \left[\frac{\partial \ln M}{\partial T_B} \right],$$

where M is the exitance of the surrounding photon background assuming an ambient temperature of T_B (300 K in the calculations). C_T is about 0.015 K^{-1} in the $8\text{--}12\text{ }\mu\text{m}$ wavelength region.

Another important figure of merit for focal plane arrays (especially arrays utilising direct injection read out circuits) is the photo generated to dark current ratio, $N_{\text{ph}}/N_{\text{th}}$ -ratio. The $N_{\text{ph}}/N_{\text{th}}$ -ratio should be as high as possible because a large N_{th} will reduce the available dynamic range of the readout circuit (the maximum charge storage capacity is limited by a integration capacitor at each pixel). Another reason for a high $N_{\text{ph}}/N_{\text{th}}$ ratio is that whereas N_{ph} can be considered independent of the detector array temperature T , N_{th} is exponentially dependent on T , which will make the array sensitive to array temperature fluctuations. NETD_{temp} and $N_{\text{ph}}/N_{\text{th}}$ ratios are listed in Table III and it is evident that decreasing T from $T = 77\text{ K}$ to $T = 70\text{ K}$ (achievable by a Stirling cooler) gives a minor decrease in NETD_{temp} and most significant, a large increase in $N_{\text{ph}}/N_{\text{th}}$ ratio. The change in grating constant from $D = 2.6\text{ }\mu\text{m}$ to $D = 2.7\text{ }\mu\text{m}$ (structure A, B and C) enhances the response at long wavelengths (although diminishes the peak absorptance slightly) of the absorptance spectrum, thus increasing the photocurrent generated in the LWIR window. Structure A reaches background limited operation (i.e. when $I_{\text{ph}} \geq I_{\text{th}}$) at $T = 77\text{ K}$ and structure D at $T = 74\text{ K}$.

5. Conclusions

We have fabricated QWIP detectors and investigated the influence of optical coupling, QW sheet doping concentration, optical response wavelength and detector temperature on QWIP performance. Peak absorption quantum efficiency (absorptance) $\eta = 87\%$ and peak detectivity $D^* = 2.8 \cdot 10^{11}$ (with $\lambda_{\text{peak}} = 8.3\text{ }\mu\text{m}$, $T = T_B = 77\text{ K}$) is demonstrated.

Table III. Calculated array performance using the measured responsivities and dark currents for $34 \times 34\text{ }\mu\text{m}$ square mesa CGW detectors. The detector temperature T is 77 K or 70 K . The calculations assume a 300 K background, optics $f\# = 2$, optical transmission 70% , maximum integration time 17 ms and a maximum charge storage capacity of 1.6×10^7 electrons. The photo current has been integrated between $8\text{ }\mu\text{m}$ to $12\text{ }\mu\text{m}$ wavelength simulating the long wavelength infrared atmospheric transmission window

Structure	Detector type	NETD _{temp} ($T = 77\text{ K}$) (mK)	NETD _{temp} ($T = 70\text{ K}$) (mK)	$N_{\text{ph}}/N_{\text{th}}$ ($T = 77\text{ K}$) (--)	$N_{\text{ph}}/N_{\text{th}}$ ($T = 70\text{ K}$) (--)
A	CGW (2.7)	12	7	0.84	7.2
B	CGW (2.7)	14	8	0.53	4.1
C	CGW (2.7)	16	8	0.39	2.6
D	CGW (2.8)	23	10	0.42	2.7

320×240 ($34 \times 34\text{ }\mu\text{m}$ detector size) array performance in terms of temporal noise equivalent temperature difference $\text{NEDT}_{\text{temp}}$ is calculated and $\text{NEDT}_{\text{temp}} = 12\text{ mK}$ (with $T = 77\text{ K}$, $T_B = 300\text{ K}$, optics $f\# = 2$ and 70% optical transmission) is obtained.

Finally, the excellent uniformity of QWIP arrays is demonstrated by measuring the dark current standard deviation to mean value ratio σ_{th} for linear arrays (1×81 and 1×128) resulting in $\sigma_{\text{th}} = 1.8\text{--}2.0\%$.

References

- Levine, B. F., Choi, K. K., Bethea, C. G., Walker, J. and Malik, R. J., *Appl. Phys. Lett.* **50**, 1092 (1987).
- Levine, B. F., *J. Appl. Phys.* **74**(8), R1 (1992) and references therein.
- Sher, A., Berding, M. A., van Schilfgaarde, M. and Chen, A.-B., *Semicon. Sci. Technol.* **6**, C59 (1991).
- Kozlowski, L. J. *et al.*, *IEEE Trans Electron. Devices* **38**, 1124 (1991).
- Levine, B. F., Bethea, C. G., Glogovsky, K. G., Stayt, J. W. and Leibenguth, R. E., "Narrow Band Gap Semiconductors", Proceedings of the NATO Workshop, Oslo, Norway, June 25–27 (1991).
- Faska, T. S. *et al.*, in "Innovative Long Wavelength Infrared Detector Workshop", Pasadena, CA, April 7–9, 1992.
- Gossen, K. W., Lyon, S. A. and Alavi, K., *Appl. Phys. Lett.* **53**, 1027 (1988).
- Hasnain, G. *et al.*, *Appl. Phys. Lett.* **54**, 2515 (1989).
- Andersson, J. Y., Lundqvist, L. and Paska, Z. F., *Appl. Phys. Lett.* **58**, 2264 (1991).
- Andersson, J. Y. and Lundqvist, L., *Appl. Phys. Lett.* **59**, 857 (1991).
- Yu, L. S., Li, S. S. and Wang, Y. H., *J. Appl. Phys.* **72**, 2105 (1992).
- Andersson, J. Y. and Lundqvist, L., *J. Appl. Phys.* **71**, 3600 (1992).
- Sarusi, G., Levine, B. F., Pearston, S. J., Bandara, K. M. S. and Leibenguth, R. E., *Appl. Phys. Lett.* **64**, 960 (1994).
- Kinch, M. A. and Yariv, A., *Appl. Phys. Lett.* **55**, 2093 (1989).
- Whitney, R. L., Cuff, K. F. and Adams, F. W., in: "Semiconductor Quantum Wells and Superlattices for Long-Wavelength Infrared Detectors" (Edited by M. O. Manasreh) (Artech House, Boston/London 1993), p. 55.

A Compact Model for the Cutoff Frequency in High Speed Bipolar Transistors

M. Andersson, P. Kuivalainen, Z. Xia, H. Pohjonen and H. Ronkainen

Technical Research Centre of Finland, Electronics, IC Group P.O. Box 11012, FIN-02044 VTT, Finland

Received May 2, 1994; accepted June 14, 1994

Abstract

A compact physical model for high speed bipolar junction transistors (BJT) in integrated rf-circuits is presented. The model, which suits both homo- and heterojunction devices, is based on the de Graaf-Kloosterman formalism for the modelling of BJTs, but adds important heterostructure device physics and incorporates also the physical properties of the SiGe material. The model implemented in APLAC circuit simulator, shows good agreement between the simulation results and measured data both for pure silicon BJTs and for SiGe-base heterojunction transistors.

1. Introduction

Recently significant progress has been achieved in the integration of the high frequency circuits by using BiCMOS or GaAs technologies. In spite of the growing use of GaAs FETs, the bipolar transistor is still the working horse in the UHF and lower microwave band circuits. Furthermore, the application of strained silicon-germanium alloys to heterojunction bipolar transistor (HBT) technology provides many additional advantages over the conventional silicon BJTs. The SiGe-base HBTs allow extreme vertical scaling without excessive base resistance for high speed performance. Also these devices can offer analog designers a very high current gain through dramatically improved Early voltages.

In analog circuit design accurate device models are needed. Recently the first model for SiGe HBTs was presented by Hong *et al.* [1]. In this paper we present a compact but physical model for high speed bipolar transistors. The model can be applied both for homo- and heterojunction devices. We have modified the original de Graaf-Kloosterman model [2] for BJTs by adding the heterostructure device physics and the physical properties of the SiGe base.

2. Model for the cutoff frequency

Cutoff frequency is naturally one of the most important parameters in the high speed bipolar transistors. It can be estimated analytically by using the charge control principle [3]

$$\frac{1}{f_T} = \frac{2\pi \sum Q}{I_C}$$

where I_C is the collector current. The total charge in the transistor is given by

$$\sum Q = Q_e + Q_b + Q_{Te} + Q_{Tc} \quad (2)$$

where Q_e is the excess hole charge in the emitter, Q_b the total base charge, and Q_{Te} and Q_{Tc} are the depletion charges of the base-emitter and base-collector junctions, respectively. So, in order to calculate the cutoff frequency we have to model the collector current I_C and the total charge in eq. (2). When generalizing the de Graaf-Kloosterman method [2] to HBTs, we have to take into account the additional electric field caused by the Ge gradient in the base region. Then the electron current density in the base is given by

$$J_n = q \langle D_n \rangle \left(\frac{2n + N_A}{n + N_A} \frac{dn}{dy} + \frac{n}{n + N_A} \frac{dN_A}{dy} + \frac{qnF_g}{k_B T} \right) \quad (3)$$

where $\langle D_n \rangle$ is the average diffusivity of electrons, n the electron concentration, F_g is the extra field due to the Ge gradient

$$F_g = \frac{1}{q} \frac{dE_g}{dy} \quad (4)$$

where E_g is the band gap.

Let us first consider the low injection case, $n \ll N_A$. Equation (3) can now be solved by integrating across the base, and we obtain

$$n(y) = n(d_{BE}) \exp(\gamma_1 y) - \frac{|J_n|}{q \langle D_n \rangle \gamma_1} [\exp(\gamma_1 y) - 1] \quad (5)$$

where

$$\gamma_1 = \frac{\eta}{W_{BM}} - \frac{qF_g}{k_B T} \quad (6)$$

and η and W_{BM} (=metallurgical base width) are the parameters which describe the doping profile for the acceptor concentration in the base, $N_A(y) = N_A^0 \exp(-\eta y/W_{BM})$. The collector current $I_C = A_e J_n$ can be solved from eq. (5) in terms of the voltage dependent minority carrier densities $n(0)$ and $n(W_b)$ at the base-emitter and base-collector junctions, respectively,

$$(1) \quad I_{CL} = q \langle D_n \rangle A_e \left[\frac{\gamma_1 \exp(\gamma_1 W_b) n(d_{BE})}{\exp(\gamma_1 W_b) - 1} - \frac{\gamma_1 n(W_b)}{\exp(\gamma_1 W_b) - 1} \right] \quad (7)$$

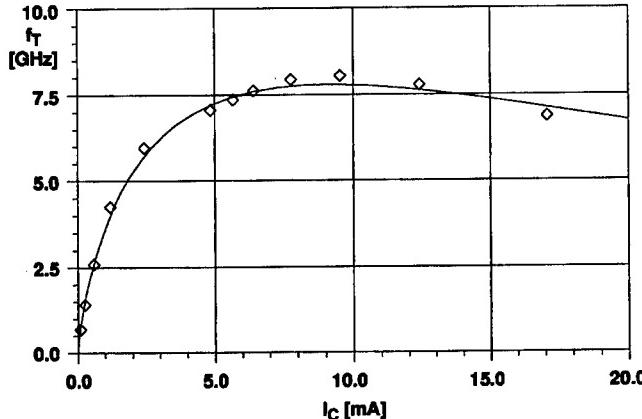


Fig. 1. Measured and modelled (solid curve) cutoff frequency *vs.* collector current in a high frequency Si BJT.

where the subscript *L* refers to low injection. The total stored base charge Q_b can also be estimated from eq. (5)

$$Q_b = qA_e \int_0^{W_b} n(y) dy \quad (8)$$

or

$$Q_b^L = \frac{qA_e}{\gamma_1} \left[\frac{\gamma_1 W_b \exp(\gamma_1 W_b)}{\exp(\gamma_1 W_b) - 1} - 1 \right] n(d_{BE}) + \frac{qA_e}{\gamma_1} \left[1 - \frac{\gamma_1 W_b}{\exp(\gamma_1 W_b) - 1} \right] n(W_b) \quad (9)$$

In the case of high injection $n \gg N_A$ the current I_{CH} and the base charge Q_b^H can be calculated as above, and then the

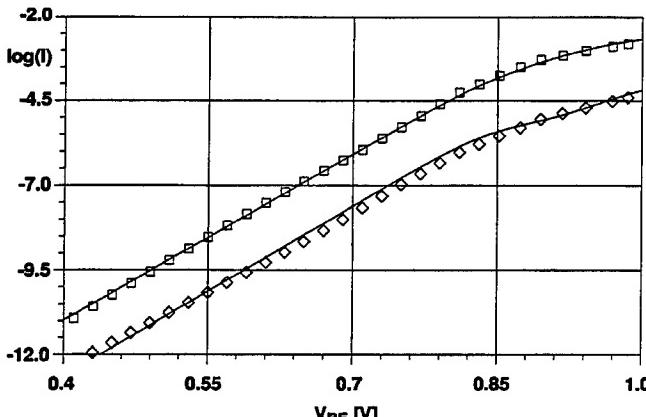


Fig. 2. Measured and modelled (solid curves) Gummel plot of a SiGe-base HBT. The upper curve is the collector current and the lower one the base current *vs.* base-emitter voltage. The experimental data was taken from [1].

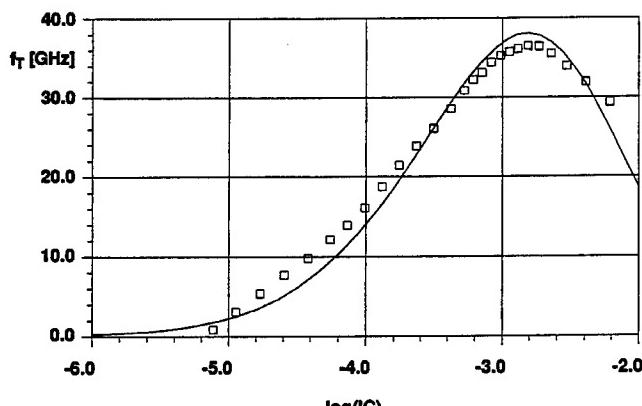


Fig. 3. Measured and modelled (solid curve) cutoff frequency *vs.* collector current in a SiGe-base HBT. The experimental data was taken from [1].

intermediate injection can be treated by using interpolation formulas [3].

The depletion charges Q_{re} and Q_{rc} at *pn*-junctions can be estimated in the usual way by integrating the depletion capacitances $C_T(V_j)$ over the junction voltages V_j . However, in the case of HBT, $C_T(V_j)$ must be modified due to the heterojunctions

$$C_T = \frac{C_0}{\left(1 - \frac{V_j}{V_{bi}}\right)^{1/2}} \quad (10)$$

where V_{bi} is the built-in voltage and

$$C_0 = A_j \left[\frac{N_D N_A \epsilon_{Si} \epsilon_{SiGe} \epsilon_0}{2V_{bi} (\epsilon_{Si} N_D + \epsilon_{SiGe} N_A)} \right]^{1/2} \quad (11)$$

Finally, the excess hole charge in the emitter can be estimated from the base current [3]

$$Q_e = \frac{W_e^2 I_B}{2 \langle D_p \rangle} \quad (12)$$

where $\langle D_p \rangle$ is the average diffusivity of the holes in the emitter, and W_e is the emitter depth. The base current in an injection-efficiency-limited BJT is governed by the properties of the emitter [3]. Therefore the base current in a homo- and heterojunction bipolar transistors are identical, and in our model we have adopted the ordinary expression for the base current [3].

When modelling the collector current dependence of f_T , the effect of the current on the depletion width of the base-collector junction has to be taken into account. Here we have adopted a simple expression given in [4].

The incorporation of germanium significantly changes the physics of the base region and the base-emitter and base-collector junctions in a SiGe-base HBT. Addition of Ge reduces the bandgap of Si, and it also modifies the energy band structure and density of states in the conduction and valence bands. In addition, charge carrier mobilities and diffusivities change due to alloy scattering and changes in the effective masses. Finally, the dielectric constant, built-in potentials and depletion widths in the *pn*-heterojunctions depend on the Ge concentration. All these dependences can be found from the literature [5–7], and they are not repeated here.

3. Model verification

The present high speed bipolar transistor model has been implemented in APLAC circuit simulator [8]. The model has been verified through comparisons between the model predictions and measured data both for Si BJTs made in our laboratory and SiGe-base HBTs made in various laboratories.

Figure 1 shows a comparison of the modelled and measured f_T *vs.* I_C for a npn Si BJT made in our laboratory. Here the base width is 270 nm, the emitter peak concentration 10^{21} cm^{-3} , the base peak concentration 10^{17} cm^{-3} , and the collector doping concentration 10^{16} cm^{-3} . Figure 1 shows a good agreement between the modelled and measured results.

Figures 2 and 3 show that in a SiGe-base HBT (8% Ge) the I-V characteristics and f_T can simultaneously be modelled

accurately, when using the present physics based device model. Here the experimental data was taken from Ref. [1]

References

1. Hong, G-B., Fossum, J. G. and Ugajin, M., IEEE Proc. International Electron Devices Meeting, 1992, pp. 577-580.
2. De Graaf, H. C. and Kloosterman, W. J., IEEE Trans. Electron Devices, **ED-32**, 2415 (1985).
3. De Graaff, H. C. and Klaassen, F. M., "Compact Transistor Modelling for Circuit Design" in Computational Microelectronics (Edited by S. Selberherr) (Springer-Verlag 1989).
4. Lee, S-G. and Fox, M., IEEE Trans. Electron Devices, **39**, 629 (1992).
5. Iyer, S. S., Patton, G. L., Stork, J. M. C., Meyerson, B. S. and Harame, D. L., IEEE Trans. Electron Devices, **36**, 2043 (1989).
6. Manku, T. and Nathan, A., IEEE Trans. Electron Devices, **12**, 704 (1991).
7. King, C. A., Hoyt, J. L. and Gibbons, J. F., IEEE Trans. Electron Devices, **36**, 2093 (1989).
8. Valtonen, M., et al., "APLAC, An Object-Oriented Analog Circuit Simulator and Design Tool", 6.2 User's Manual and Reference Manual, (Helsinki University of Technology, Circuit Theory Laboratory and Nokia Research Center, Hardware Design Technology), March 1994.

Extrinsic Versus Intrinsic Models for FETs

T. Ytterdal†*, T. A. Fjeldly† and K. Lee*

† Department of Physical Electronics, Norwegian Institute of Technology, University of Trondheim, N-7034 Trondheim, Norway

* Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Taejon, Korea

Received April 28, 1994; accepted June 14, 1994

Abstract

The importance of a proper inclusion of parasitic source and drain resistances in various FET device models used in circuit simulation with SPICE is pointed out. Although a significant reduction in simulation time can be achieved using so-called extrinsic FET models, some problems are encountered in cases where gate leakage current is present and in simulating transients. Moreover, an intrinsic model with parasitics is more compatible with high frequency small signal equivalent circuits.

1. Introduction

As integrated circuit technology advances and the channel length of FETs are steadily reduced, the importance of the source and drain parasitic resistances, R_s and R_d , become increasingly important in the overall operation of the devices. This is due to the fact that in scaled down devices, the value of the channel resistance becomes comparable to the parasitic resistances. In the equivalent circuits used for modeling such devices, it has been common to either neglect R_s and R_d altogether (intrinsic model) or to combine the intrinsic model with additional nodes to account for the parasitics.

However, with the introduction of new analytical techniques, it became possible to include R_s and R_d directly into the expression for the drain current, resulting in so-called extrinsic models [1, 2]. An immediate benefit of extrinsic models is a reduction in the node count which is favorable in terms of circuit simulation time. However, the extrinsic models may also have drawbacks in modeling certain FETs such as MESFETs and HEMTs operating in regimes where the gate leakage current is significant.

In this paper, we compare extrinsic and intrinsic models in terms of computing speed and accuracy. We also point out problems with extrinsic models encountered in cases where gate leakage current is present and in simulating transients.

2. Ways to include parasitic resistances

We have basically the following three approaches for including the effects of parasitic source and drain resistances into FET equivalent circuits:

A. No explicit parasitics

The simplest approach, still being used in circuit simulation today, is to use an intrinsic model with no explicit addition of R_s and R_d . In this case, the intrinsic model parameters are extracted directly from the measured $I-V$ characteristics,

but effects of R_s and R_d will, to some extent, implicitly be contained in the remaining parameters. One possibility is to include the effects of the parasitics in a mobility reduction parameter [3].

For MOSFETs, we may be able to faithfully reproduce measured DC characteristics this way. However, for MESFETs this type of model yields an unrealistic increase in the gate leakage current with positive gate voltage.

Furthermore, we can expect problem in simulating transient and small-signal behavior of all types of FETs since important time constants are related to the presence of the parasitic resistances.

B. Intrinsic models with external parasitics

This approach is the most widely used today. The procedure for obtaining a good fit to measured $I-V$ characteristics is to first extract geometrical information and then determine R_s and R_d . This information is then used to derive the intrinsic characteristics of the device, from which the parameters of the intrinsic device can be extracted. With this technique, the equivalent circuit will contain the parasitic resistances at the expense of additional nodes in the circuit.

C. Extrinsic models

Here, the parasitic resistances are included as an integral part of the model expressions and treated on equal footing with the intrinsic model parameters. Hence, there is no need to add external resistances to the equivalent circuit, and the corresponding nodes are saved. The mathematical description of an extrinsic model will, of course, be more complex than for the intrinsic model, but none-the-less, we should expect a savings in terms of simulation time owing to the reduced node count.

To illustrate this, we used AIM-Spice [4, 5] to calculate the operating point of MOSFET inverter chains using intrinsic (with external resistances) and extrinsic versions of the model discussed in [1]. The simulation time was monitored for a range of inverter stages in the chain, and the results are shown in Fig. 1.

As can be seen from the figure, we consistently find a reduction of about 25% in the simulation time for the extrinsic model compared to the intrinsic one with external R_s and R_d . A slight decrease in this ratio of the simulation times with complexity probably reflects the fact that the system matrix for the intrinsic model increases faster than that of the extrinsic model.

The extrinsic model approach is similar to that discussed above for the intrinsic model with no external parasitics. Unfortunately, it also has similarities in terms of the problems pointed out with respect to gate leakage current (when

* Corresponding author.

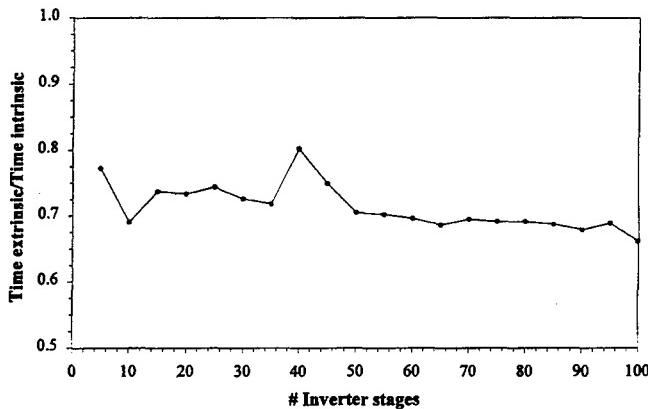


Fig. 1. Ratio of simulation times with AIM-Spice using extrinsic and intrinsic models versus number of MOSFET inverter stages.

applicable) and transient behavior. To illustrate the latter, we performed transient simulations of a CMOS circuit consisting of two inverters in series, using the same models as in the previous example. As expected, the $I-V$ characteristics with both sets of models were identical. But noticeable differences could be observed in the *on-to-off* transient for the circuit in two cases, as indicated in Fig. 2.

Obviously, the extrinsic version produces a more pronounced peak in the output at the start of the pull-down

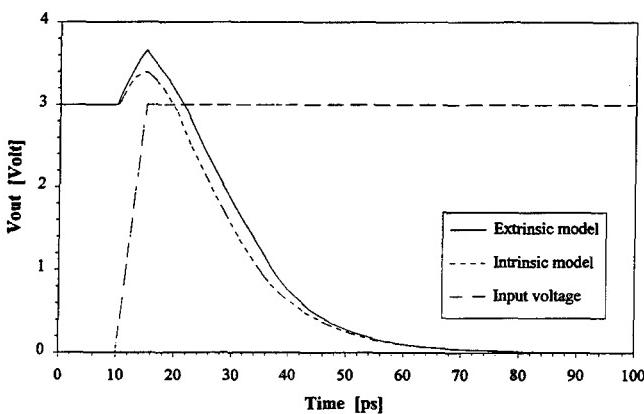


Fig. 2. Comparison of transient analyses with AIM-Spice for a two stage inverter circuit using extrinsic and intrinsic models. Council of Norway.

phase, and the fall time is slightly increased compared to the intrinsic version.

Finally it should be pointed out that an intrinsic model with explicit parasitic is more compatible with typical small signal equivalent circuits (see for example [6]).

3. Conclusion

With the introduction of new analytical techniques, extrinsic FET device models for circuit analysis have emerged. The main advantages compared to intrinsic models are: simplified parameter extraction, reduced node count and reduced simulation time in SPICE. However, some drawbacks are noted, including problems with accurate modeling of gate leakage current in devices such as MESFETs and HEMTs, and in the simulation of transient and small-signal behavior in all types of FETs. Moreover, proper inclusion of the parasitics is also very important in order to extract the effect of contact technologies on circuit behavior. In cases where these issues are important, the use of intrinsic device models with extrinsic parasitic resistances give higher accuracy.

Acknowledgement

The work was supported by Hyundai Electronics Industries Co., Ltd., NATO Scientific Affairs Division and the Research Council of Norway.

References

1. Fjeldly, T. A., Shur, M., Lee, K. and Ytterdal, T., "Proceedings of ISDRS'91" (Charlottesville, Virginia 1991), p. 407.
2. Chow, H. C. and Feng, W. S., IEEE Trans. Electron Devices TED-39, 2626 (1992).
3. Suciu, P. I. and Johnson, R. L., IEEE Trans. Electron Devices TED-27, 1846 (1980).
4. Ytterdal, T., Lee, K., Shur, M. and Fjeldly, T. A., "Proceedings of ISDRS'91" (Charlottesville, Virginia 1991), p. 481.
5. Lee, K., Shur, M., Fjeldly, T. A. and Ytterdal, T., "Semiconductor Device Modeling for VLSI" (Prentice Hall, New Jersey 1993).
6. Ladbrook, P. H., "MMIC Design with GaAs FETs and HEMTs" (Artech House, 1988).

The Effect of Using Different Transport Models in Computer Simulations of the Permeable-Base Transistor

Hans-Erik Nilsson,^{1,2} Ulf Sannemo,² Ants Koel,^{3,4} Ferenc Masszi³ and C. Sture Petersson¹

¹ Department of Solid State Electronics, The Royal Institute of Technology, Electrum 229, S-164 40 Kista, Sweden

² School of Science and Engineering, Mid-Sweden University S-851 70 Sundsvall, Sweden

³ Scanner Lab, Electronics Department, Institute of Technology, Uppsala University, P.O. Box 534, S-751 21 Uppsala, Sweden

⁴ Electronics Department, Tallinn Technical University, Estonia

Received June 6, 1994; accepted June 14, 1994

Abstract

The PBT is a short channel device and hot electron effects are expected to be important and the transport model used in the simulation is critical. The present work compares the effect of different transport models on the operation of a submicron PBT. The transport model used are the ordinary drift-diffusion model and the hydrodynamic models as they are implemented in MEDICI Ver. 1.1. In submicron devices the transport parameters become both device and bias dependent. The transport parameters are directly related to the distribution function and should be extracted from the solution of Boltzmann's Transport Equation (BTE). The most popular and effective way of solving the BTE is Monte Carlo simulation. In this work we have extracted transport parameters from Monte Carlo simulation using one dimensional boundary for particles and a fixed electric field extracted from two-dimensional simulation in MEDICI. The Monte Carlo simulation, parameter extraction and MEDICI simulation have been repeated until the change in electric field between iterations could be neglected. Both $I-V$ and f_T results have been analysed. It is clear that the hydrodynamic model gives higher current levels and higher f_T than the drift-diffusion model.

1. Introduction

Computer simulation for predicting device behaviour is used almost as an inevitable tool today before starting up any processing. A computer model of any art always consists of two parts: first, through simplifications and approximations, a physical model of the reality is created, and then this model is tried to be solved with adequate mathematical methods. In the case of semiconductor modelling, the general physical model is based on estimating the charge/potential relation (described by Poisson's equation, originating from Maxwell's equations), and the carrier transport model. The transport model is originating from the general Boltzmann Transport Equation, which is an extremely complicated equation from the mathematical (i.e. from the future solution) point of view. The only question is, how long one can proceed in the simplification of the physical abstraction, while preserving the important features, and at the same time making a mathematical solution still possible.

Van Roosbroeck [1] was the first to suggest the following set of differential equations (all notations have their usual meaning)

Poisson's equation:

$$\operatorname{div}(\epsilon \operatorname{grad} \phi) = -\rho \quad (1)$$

Continuity equations:

$$\operatorname{div} \bar{J}_n - q(\partial n / \partial t) = qU \quad (2)$$

$$\operatorname{div} \bar{J}_p + q(\partial p / \partial t) = -qU \quad (3)$$

Transport equations:

$$\bar{J}_n = q\mu_n n \bar{E} + qD_n \operatorname{grad}(n) \quad (4)$$

$$\bar{J}_p = q\mu_p p \bar{E} - qD_p \operatorname{grad}(p) \quad (5)$$

This model is based on the relaxation-time approximation during the solution of the Boltzmann equation, and is generally called the drift-diffusion model, because of its current terms. However, by shrinking dimensions, the inadequacy of this model becomes more and more evident: the continuum view of the electron population is becoming questionable, and this asks for a more elaborate transport model.

The hydrodynamic model was first suggested by R. Stratton [2]. Later on several authors contributed to the particular version that is used in this paper [3–5]. In this model the Boltzmann equation

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + u \operatorname{grad}_r f \pm \frac{2\pi q}{h} \bar{E} \operatorname{grad}_k f = \left(\frac{\partial f}{\partial t} \right)_{\text{collisions}} \quad (6)$$

is solved under the following assumptions [6]

- The temperature tensor can be expressed as a scalar quantity
- The heat flux can be expressed as proportional to the same gradient
- The collision integrals can be expressed via momentum- and energy-relaxation time approximations. (The relaxation time constants are determined from Monte Carlo simulations.)

The following set of transport equations called the hydrodynamic model (HD) is formed

$$\bar{J}_n = q\mu_n \left[n \bar{E} + \nabla \left(\frac{k_B T_n}{q} n \right) \right] \quad (7)$$

$$\bar{J}_p = q\mu_p \left[p \bar{E} - \nabla \left(\frac{k_B T_p}{q} p \right) \right] \quad (8)$$

$$\nabla \bar{S}_n = \frac{1}{q} \bar{J}_n \bar{E} - \frac{3}{2} \left[\frac{nk_B(T_n - T_0)}{q\tau_{\text{rel. } n}} + \frac{\partial \left(\frac{nk_B T_n}{q} \right)}{\partial t} \right] + E_g G_n^H \quad (9)$$

$$\nabla \bar{S}_p = \frac{1}{q} \bar{J}_p \bar{E} - \frac{3}{2} \left[\frac{pk_B(T_p - T_0)}{q\tau_{\text{rel. } p}} + \frac{\partial \left(\frac{nk_B T_p}{q} \right)}{\partial t} \right] + E_g G_p^H \quad (10)$$

$$\bar{S}_n = -\frac{5k_B T_n}{2q} \left[\frac{\bar{J}_n}{q} + \kappa_n \mu_n n \nabla \left(\frac{k_B T_n}{q} \right) \right] \quad (11)$$

$$\bar{S}_p = \frac{5k_B T_p}{2q} \left[\frac{\bar{J}_p}{q} + \kappa_p \mu_p p \nabla \left(\frac{k_B T_p}{q} \right) \right] \quad (12)$$

where the new quantities are: $\bar{S}_{n,p}$ = energy flow density, E_g = band-gap, $G_{n,p}^H$ = impact ionization coefficient, $\kappa_{n,p}$ = thermal conductivity coefficient, $\mu_{n,p}$ is a function of $T_{n,p}$.

Another similar version uses the following equations instead of (7) and (8):

$$\bar{J}_n = q\mu_n \left[n\bar{E} + \nabla \left(\frac{k_B T_n}{q} \right) \right] + n \frac{k_B T_n}{q} \frac{\partial \mu_n}{\partial T_n} \nabla(T_n) \quad (13)$$

$$\bar{J}_p = q\mu_p \left[p\bar{E} - \nabla \left(\frac{k_B T_p}{q} \right) \right] - p \frac{k_B T_p}{q} \frac{\partial \mu_p}{\partial T_p} \nabla(T_p) \quad (14)$$

In this paper we refer to this model as the energy transport model (ET). The largest difference, compared to the drift-diffusion model, is the calculation of the carrier heating, which makes it possible to include hot electrons, velocity overshoot and other non-static effects in the simulation.

2. Known results

The Permeable-Base Transistor (PBT) has gained a certain interest in the past years. The main interest in PBTs is to use them as microwave active components. During the years, there have been many optimistic predictions published about the transistor's potential in high frequency applications. Estimated maximum frequency of oscillation has been found to be over 1000 GHz for GaAs PBT [7].

Monte Carlo calculations in [8] indicates unity current gain frequency f_T in the range of 120 GHz for a Si PBT. Recent progress in process technology has made it possible to produce high performance PBTs. Some examples of results are a GaAs PBT with unity power gain frequency f_{max} of 220 GHz [9] and a Si PBT with a unity current gain frequency just above 25 GHz [10]. We have chosen to focus our interest exclusively on Si PBTs.

Figure 1 shows a section of an overgrown PBT used in this paper. Depending on the depletion layer width at the Schottky contact, the PBT can work as a vertical MESFET or actually as PBT. In the PBT mode, the channel is depleted and the net current is mainly a diffusion current, which makes it more related to a bipolar transistor than to

a MESFET [7]. Unlike ordinary MESFETs the channel width and length are strong functions of the potential distribution in the channel. For PBT operation the line spaces in the grid have to be very small, less than two times the depletion layer width. For CoSi₂ and $N_D = 10^{17} \text{ cm}^{-3}$ the line space should be under 0.15 micron. In spite of tremendous progress in process technology, it is still very difficult to fabricate lines and line spaces under 200 nm. Therefore, we have focused our investigation on PBTs in MESFET operation.

As the most important feature of the PBT is its AC behaviour, it is obvious that the aim of a simulation should be to determine the frequency limit. The unity current gain frequency f_T can be calculated as [11]:

$$f_T = \frac{g_m}{2\pi C_T} = \frac{1}{2\pi} \frac{\Delta I_{ds}}{\Delta Q_t} \quad (15)$$

ΔI_{ds} , ΔQ_t = current and charge difference between two different V_{gs} bias voltages (steady state).

3. Present work

Simulations of the PBT in [8] and [12] indicate large differences between the drift-diffusion model and Monte Carlo simulation. The Monte Carlo simulation of a GaAs PBT in [12] shows a decrease in the gate capacitance and an increase in transconductance compared to the drift-diffusion model. For a GaAs PBT with uniform doping f_T increases with 60% compared to the drift-diffusion model. Monte Carlo simulations is time consuming and not practical for use for optimization of a device. The hydrodynamic model is much more suitable for this purpose and commercial software packages like MEDICI are available which provides state of the art hydrodynamic models, models for Schottky-barriers and mobilities [13].

Figure 1 shows the structure we used in all our simulations. The unity current gain frequency f_T should increase as device dimensions shrink. The channel in a PBT can be made shorter if the distance between drain and source is reduced. Ohshima *et al.* [8] has presented a comparison between a drift-diffusion model and Monte Carlo calculations where the source to drain distance l_{ds} was changed from 200 to 600 nm. The grid spacing was 200 nm and the grid width was 200 nm. The simulation result presented in [8] shows a drastic increase in f_T for short devices. Our intention in this work was to see whether the HD or the ET model can be used for device optimization taking into account the hot electron effects. We used the same test setup as the one used in [8]. The unity current gain frequency f_T was extracted for $V_{gs} = 0 \text{ V}$ and $V_{ds} = 3 \text{ V}$. We have used the parallel field-dependent mobility for the drift-diffusion model, temperature dependent mobility for the hydrodynamic model, and the concentration dependent mobility for low electric fields [14].

Parallel Field-Dependent mobility

$$\mu_{n,p}^{\text{low}} = \left[\frac{\mu_{n,p}^{\text{low}} E_{n,p}^{\parallel}}{1 + \left(\frac{\mu_{n,p}^{\text{low}} E_{n,p}^{\parallel}}{v_{n,p}^{\text{sat}}} \right)^{\beta}} \right]^{1/\beta},$$

$$\mu_{n,p}^{\text{low}} = \text{low field mobility} \quad (16)$$

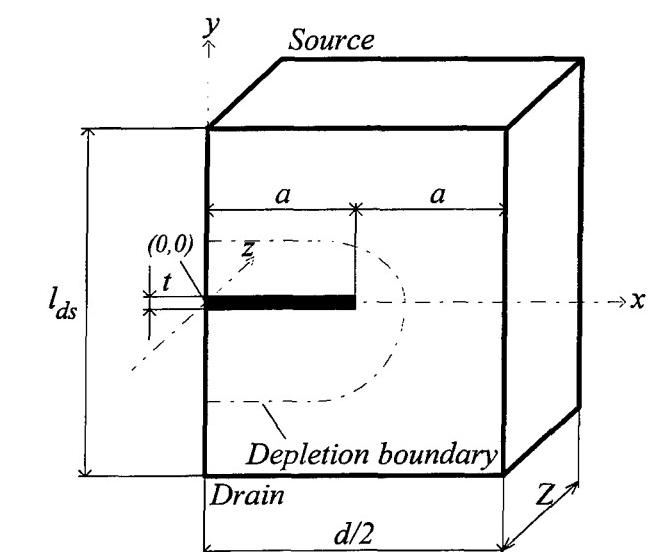


Fig. 1. PBT structure used in this paper. The period length of the grid is $d = 4a$.

Temperature Dependent Mobility

$$\mu_{n,p} = \frac{\mu_{n,p}^{\text{low}}}{\left[1 + \left(\frac{3\mu_{n,p}^{\text{low}} k_B}{2q(v_{\text{sat}})^2 T_{n,p}^{\text{rel}}} (T_{n,p} - T_{\text{lattice}}) \right)^{1/\beta} \right]^\beta} \quad (17)$$

In the first comparison we have used MEDICI's default parameters for the energy relaxation time and velocity saturation which is 0.2 psec and 1×10^7 cm/s. The concentration dependent mobility model in MEDICI uses a look-up table for the relationship between the mobility and the impurity concentration [11]. In Fig. 2 we have plotted f_T as a function of device length for both the drift-diffusion (DD) and the hydrodynamic model (HD) using the default transport parameters from MEDICI. The hydrodynamic model predicts a higher f_T than the drift-diffusion model but the increase is far from the predictions made by Monte Carlo simulation in [8] and [12]. Our drift-diffusion simulation does not agree with the drift-diffusion simulation in [8]. This is not surprising since in [8] the "drift velocity was kept constant (1×10^7 cm/s)", which is the same as using an infinite low field mobility in (17). MEDICI's more realistic model gives two times higher f_T for long devices compared to the result in [8] but much lower f_T for shorter devices. The constant velocity model is not correct and a better mobility model has to be used in order to get the correct field distribution in the channel. The Monte Carlo results in [8] and [12] give f_T about two times higher than the drift-diffusion prediction. The results from the HD model does not show this type of increase in f_T .

In order to estimate the hot electron effect in the channel in a more exact way we have simulated the particle dynamics in a full band Monte Carlo program with one dimensional particle boundaries using an electric field extracted from the hydrodynamic solution in MEDICI. We have used two different devices in our simulations (1) 400 nm in source to drain distance, 200 nm in grid spacing, 200 nm in grid width, (2) 200 nm in source to drain distance, 150 nm grid spacing, 150 nm grid width. The electric field was extracted along the path of maximum current in the device which is along a line segment from source to drain through the centre of the grid spacing. These simulations showed that the default energy relaxation time used was too low since the electron temperature was underestimated compared to the Monte Carlo simulation.

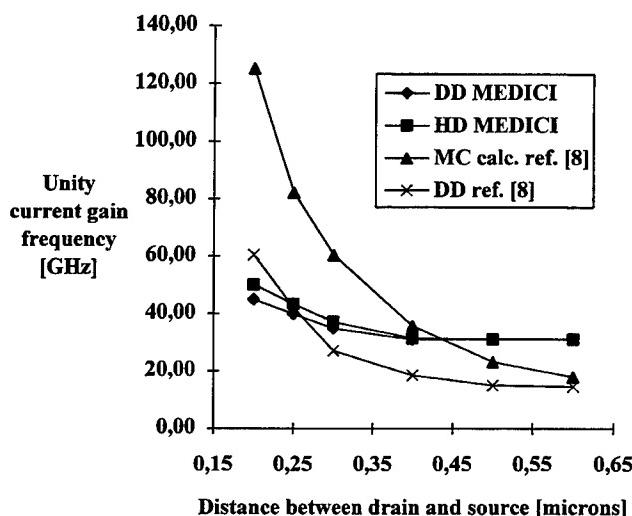


Fig. 2. Unity current gain frequency as a function of device length.

The transport parameters used in the HD and ET models are assumed to be independent of electric field and position within the device. This is not correct for small devices where the transport parameters are functions of device dimensions and voltage biasing and should be extracted from the solution of Boltzmann's Transport Equation. However, if the distribution function is known we have all information and no further simulation is needed. We have used Monte Carlo simulation with one-dimensional particle boundaries to find an approximation of the distribution function which can be used to extract better transport parameters than the default parameters in MEDICI. The Monte Carlo program is based on a full band code form University of Illinois [15]. The program has been upgraded to include many particle simulation, constant time technique for the free flight generation and a fourth and fifth order Runge-Kutta algorithm for solving the equation of motion. Further details on the scattering models used can be found in [15]. The Monte Carlo simulation, parameter extraction and MEDICI simulation were repeated until the change in electric field could be neglected. The number of iterations is small since the electric field is rather insensitive to changes in the transport parameters. The velocity saturation was extracted from bulk simulation and found to be 1.15×10^7 cm/s. The velocity vs. electric field extracted from Monte Carlo is different for increasing and decreasing field and a low field mobility is difficult to extract in an accurate way. We have used a low field mobility of $675 \text{ cm}^2/\text{Vs}$ which is in agreement with the concentration mobility model in MEDICI. This value gives a velocity vs. electric field curve that is lower than the velocity obtained from Monte Carlo simulation for increasing field but higher than the velocity for the decreasing field. The energy relaxation time has been chosen so that the electron temperature fits the result obtained from the Monte Carlo simulation. Using the method described above transport parameters was extracted for the mobility model (18) both for $\beta = 1$ and $\beta = 2$. A comparison between the velocity and electron temperature obtained from Monte Carlo, HD and ET models can be seen in Fig. 3 and Fig. 5. The ET model gives the best fit in velocity and also the highest current and the highest f_T as can be seen in Fig. 4 and Fig. 6. The simulation results shows that f_T is larger for HD and ET compared to the drift-diffusion model, but still far from the values predicted in [8] and [12]. The ET model and the HD model in MEDICI gives higher transconductance but the capacitance at the gate will also increase. This increase in the capacitance is due to the change in the potential distribution in the channel, which will change the depletion boundary and the number of charges in the depletion region.

4. Discussion

We expect the ET and the HD model to be more accurate than the drift-diffusion model because of the better grade of approximation in the transport model. However, the difference is not that large (although it seems to be more and more significant with smaller dimensions). The Monte Carlo results obtained in [12] is for GaAs and a direct comparison with silicon cannot be done. The velocity overshoot is much larger in this material and the velocity peak is wider.

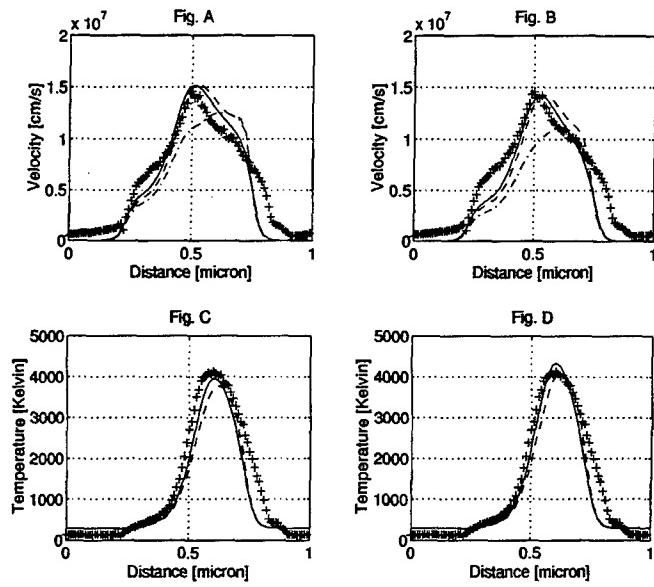


Fig. 3. (a) Electron velocity for $\beta = 2$. (b) Electron velocity for $\beta = 1$. (c) Electron temperature for $\beta = 2$. (d) Electron temperature for $\beta = 1$. All data is extracted along the channel at $x = 2a$. (+) Monte Carlo, (solid line) ET model, (dashed line) HD model. Device dimensions are $2a = 0.2$ micron, $l_{ds} = 1.0$ micron (400 nm between high doping regions gives an effective $l_{ds} = 400$ nm), $t = 40$ nm.

A full self consistent Monte Carlo simulation of a silicon PBT with a detailed description of the models used cannot be found in the literature. The Monte Carlo calculations presented in [8] does not include any detailed information about the physical models used. The f_T for the Monte Carlo calculation has the same limit value as the drift-diffusion case for large source to drain distances in spite of a neglected low field mobility in the drift-diffusion model. For these dimensions a correct full Monte Carlo simulation is expected to give a result that is consistent with a drift-diffusion model which includes a reasonable low field mobility.

5. Conclusion

The effect of using different transport models in computer simulation of the PBT has been discussed. It has been

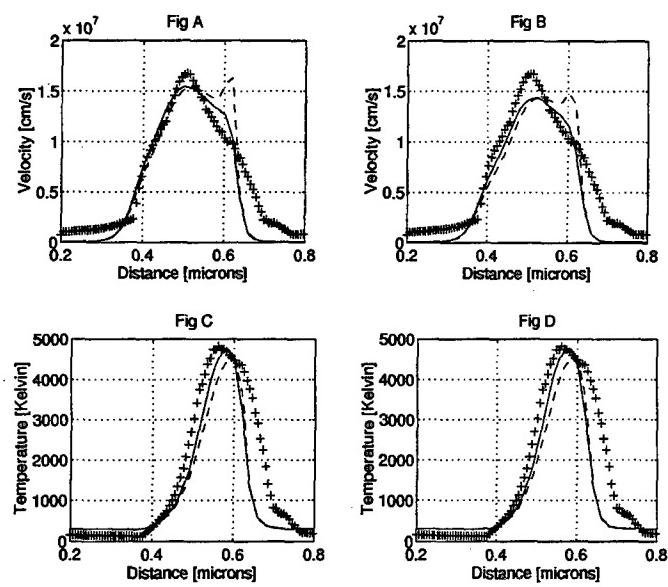


Fig. 5. (a) Electron velocity for $\beta = 2$. (b) Electron velocity for $\beta = 1$. (c) Electron temperature for $\beta = 2$. (d) Electron temperature for $\beta = 1$. All data is extracted along the channel at $x = 2a$. (+) Monte Carlo, (solid line) ET model, (dashed line) HD model. Device dimensions are $2a = 0.15$ micron, $l_{ds} = 1.0$ micron (0.4 micron between high doping regions gives an effective $l_{ds} = 0.4$ microns), $t = 40$ nm.

shown that the hot electron effect will have a strong influence on the drain current. The hydrodynamic model HD and the energy transport model ET does not predict a large increase in f_T compared to the ordinary drift-diffusion model even if the transport parameter gives a velocity overshoot effect in agreement with Monte Carlo simulations using the same electric field. The ET model provides better agreement with the Monte Carlo model and gives the highest current levels and f_T values. In the simulations of small structures the HD model gives a small velocity peak where the electric field drops. This has not been seen in the Monte Carlo model or in the ET model. In spite of this extra velocity peak the transconductance is lower than for the ET model. A comparison between velocity and electron temperatures obtained from Monte Carlo simulations and the HD and ET model shows that the energy relaxation time should be considered as a device dependent parameter. For device

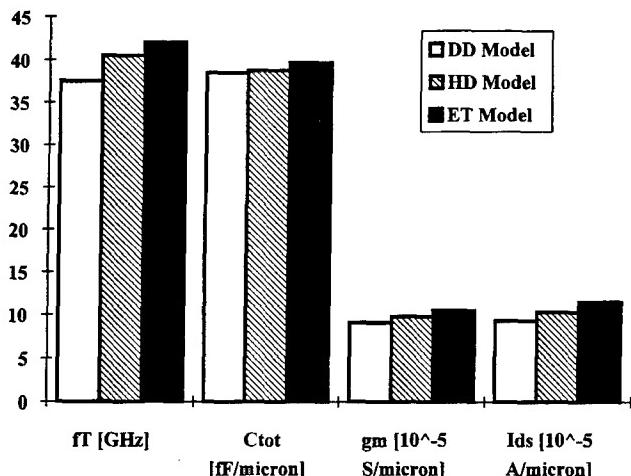


Fig. 4. f_T , g_m , C_{tot} and I_{ds} for different transport models ($\beta = 2$). Device dimensions are $2a = 0.2$ micron, $l_{ds} = 1.0$ micron (400 nm between high doping regions gives an effective $l_{ds} = 400$ nm), $t = 40$ nm. $V_{ds} = 3.0$ V, $V_{gs} = 0.0$ V.

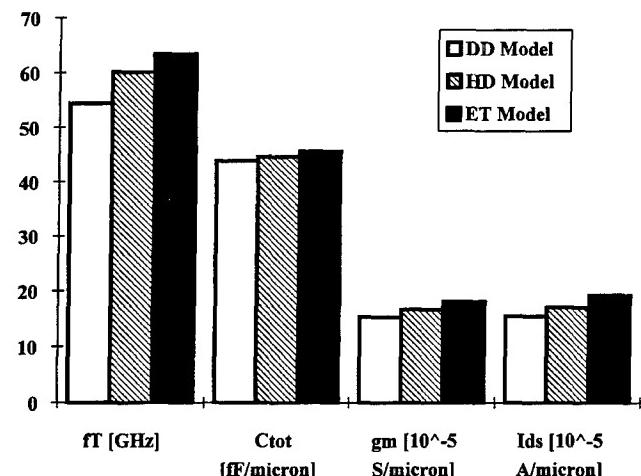


Fig. 6. f_T , g_m , C_{tot} and I_{ds} for different transport models ($\beta = 2$). Device dimensions are $2a = 0.15$ micron, $l_{ds} = 1.0$ micron (200 nm between high doping regions gives an effective $l_{ds} = 200$ nm), $t = 40$ nm. $V_{ds} = 3.0$ V, $V_{gs} = 0.0$ V.

dimensions under 200 nm the HD and ET model have problems to fit to the Monte Carlo model and a Monte Carlo simulation including a self-consistent solution of Poisson's equation has to be used in order to evaluate the high frequency performance in an accurate way.

References

1. Van Roosbroeck, W. V., Bell Syst. Tech. J., **29**, 560 (1950).
2. Stratton, R., Phys. Rev., **126**(6), 2002 (1962).
3. Blotekaer, K., IEEE Trans. on Elec. Dev., **ED-17**, 38 (1970).
4. Baccarani, G. and Wordeman, M., Solid-State Electronics, **28**(4), 407 (1985).
5. Forghieri, A. et al., IEEE Trans. on Computer-Aided Design, **7**(2), 231 (1988).
6. Baccarani, G., Rudan, M., Guerrieri, R. and Ciampolini, P., "Process and Device Modeling" (Edited by W. L. Engl) (Elsevier Science Publisher B. V., North-Holland 1986), ch. 4, p. 107.
7. Bolzer, C. O. and Alley, C. D., IEEE Trans. on Elec. Dev., **ED-27**(6) (1980).
8. Ohshima, T., Nakamura, N., Nakagawa, K., Yamaguchi, K. and Miyao, M., IEDM 91, 33 (1991).
9. Hollis, M. A., Nichols, K. B., Murphy, R. A. and Bozler, C. O., SPIE, Advanced Processing of Semiconductor Devices, **797** (1987).
10. Gruhle, A. et al., Microelectronic Engineering, **15**, 27 (1991).
11. Rathman, D. D., IEEE Trans. on Elec. Dev. **ED-37**(9), 2090 (1990).
12. Hwang, C-G., Navon, D. H. and Tang, T-W., IEEE Trans. on Elec. Dev. **ED-34**(2) (1987).
13. Technology Modeling Associates, Inc., "Reference Manual of MEDICI, Ver. 1.1", **1**, 2-20, 2-67, 2-68 (1993).
14. Thornber, K. K., J. Appl. Phys. **51**, 2127 (1980).
15. Shichijo, H., Tang, J. Y., Bude, J. and Yoder, D., "Monte Carlo Device Simulation: Full Band and Beyond" (Edited by K. Hess) (Kluwer Academic Publishers 1991), ch. 10, p. 285.

Monte Carlo Simulation of Hot Carrier Noise in Short n^+nn^+ Diodes

V. Gružinskis, E. Starikov and P. Shiktorov

Semiconductor Physics Institute, A. Goštauto 11, 2600 Vilnius, Lithuania

Received April 28, 1994; accepted June 14, 1994

Abstract

The Monte Carlo method is applied to calculate the current and voltage noise in near micron n^+nn^+ InP diodes. Quite different behavior of the correlation functions and spectral densities of fluctuating macroscopic quantities is observed under the voltage and current driven operations. Under the constant voltage operation the time dependence of the current fluctuation correlation function exhibits damped oscillations at the transit-time and plasma frequencies. This results in appearance of two spikes in the current noise spectrum at corresponding frequencies. The transit-time oscillations and the corresponding noise are shown to be caused by the spontaneous formation of electron accumulation layers due to the negative differential resistance connected with the combined action of the velocity overshoot and Gunn-effects. In the contrast, the voltage noise spectrum is found to have a regular Lorentzian shape under the constant current operation. The observed features of the current and voltage noise spectra are shown to be in a good agreement with the frequency dependence of the small-signal admittance and impedance of the diode.

1. Introduction

Nonequilibrium noise is one of the important features of hot carriers and plays a rather essential role in characterization of carrier transport in bulk semiconductors and semiconductor devices [1–6]. For comprehensive theoretical analysis of hot-carrier noise in the time- and frequency-domains the use is usually made of the auto- and cross-correlation functions of various fluctuating quantities and their spectral densities [7–10]. In bulk semiconductors the time and frequency behavior of the velocity and energy fluctuations reflects both the dynamic and relaxation processes occurring in the momentum space and, as such, can be used for a detailed investigation of the physical processes inherent in the hot-carrier homogeneous steady-state itself providing useful information about various relaxation rates, internal cyclic processes, etc. [8–10]. In small semiconductor devices the velocity and energy fluctuations becomes strongly coupled with fluctuations of the local carrier concentration and resulting electric field, thus, leading to appearance of an additional noise connected with the plasma and transit-time effects. For the device performance these noise components are of great practical importance since they determine the lower limit of a device sensitivity, indicate the onset of generation processes, etc. [3–6]. The aim of this work is to provide a theoretical investigation of the hot-carrier noise in nearmicron n^+nn^+ InP diodes under biasing conditions for which the microwave power generation associated with velocity overshoot and electron valley-transfer is possible.

2. Procedure

The Monte Carlo Particle (MCP) method is the most appropriate technique for hot-carrier noise investigations

since it allows the appropriate correlation functions to be calculated in a natural way, by using a time-averaging over a multi-particle history simulated during a sufficiently long time interval. In the present work MCP method is used to calculate both the current and voltage noise spectral densities. These noise operations correspond to idealized conditions when either the voltage U_d applied to the diode or the total current per unit area j_{tot} flowing through the diode are kept constant in time [7, 9]. In the former case the fluctuating quantity is the conduction current per unit area $j_{cond}(t)$ which is given by the sum of instantaneous velocities of all carriers inside the device [9]. Since U_d is kept constant, the conduction current coincides with the total current in this case. Under the constant current operation the fluctuating quantity is the voltage drop between the diode terminals. Its variation in time is described by a differential equation:

$$\frac{d}{dt} U_d(t) = \frac{l}{\epsilon_0 \epsilon_r} [j_{tot} - j_{cond}(t)] \quad (1)$$

where ϵ_0 is the vacuum permittivity and ϵ_r the static dielectric constant, l the diode length. The total current j_{tot} is kept constant in time, so the voltage fluctuations results from fluctuation of the conduction current. Therefore, the microscopic fluctuations of the instantaneous velocities of single carriers are responsible for both the current and voltage noise which is detected outside the diode in an external circuit. It should be stressed that from the physical point of view the above operations correspond to the idealized conditions of the diode performance in the external circuit. The current noise operation can be realized when the diode resistance R_d is much greater than the external load resistance R . The conditions for the voltage noise operation are fulfilled in the opposite case. Such an interpretation allows for the diode noise to be described in terms of the equivalent sources of either the current or voltage noise [3–5]. Thus, investigation of the current and voltage noise corresponds to determination of the main temporal and spectral characteristics of the equivalent sources. To investigate the diode noise in these cases the correlation function of the fluctuations of the corresponding quantity $Q(t)$ (i.e. $j(t)$ and $U_d(t)$ for the current and voltage noise, respectively) is calculated as:

$$C_Q(t) = \frac{1}{T} \int_0^T \delta Q(t') \delta Q(t' + t) dt' \quad (2)$$

where T is the averaging time, $\delta Q(t')$ the fluctuation of $Q(t')$, t the delay time. Then the noise spectral density $S_Q(v)$ at the frequency v is calculated in the standard way as:

$$S_Q(v) = 4 \int_0^\infty C_Q(t) \cos(2\pi v t) dt \quad (3)$$

It should be underlined that the spectral densities of the current and voltage fluctuations are mutually connected through the small-signal impedance of the diode normalized to the unit cross-section [5]:

$$S_U(v) = |Z(v)|^2 S_j(v) \quad (4)$$

As it follows from eq. (4) a description of the noise induced in the external circuit by a whole device involves into consideration also its small-signal characteristics. Therefore, to clarify the physical origin of the hot carrier noise for the same diodes we have also calculated the frequency dependence of the small-signal impedance $Z(f)$ by using a closed hydrodynamic (HD) approach. The details of the MCP and HD techniques used in our calculations can be found elsewhere [11].

3. Results

Calculations are performed for the case of n^+nn^+ InP diode at room temperature with the doping levels of $n = 2 \times 10^{16}$ and $n^+ = 10^{18} \text{ cm}^{-3}$. The cathode, n -region and anode lengths are respectively 0.05, 1.0 and 0.1 μm . Abrupt homojunctions are assumed. Similar structures are widely used in modern microwave power generators [12, 13]. As it has been shown by the hydrodynamic modelling [14, 15] the dynamic negative differential resistivity (NDR) of this diode appears above the threshold voltage $U_{th} = 1.5 \text{ V}$ and covers the frequency range $v = 100\text{--}250 \text{ GHz}$ where the real part of the small-signal impedance, $\text{Re } Z(v)$, is negative. Under voltage-driven operation this diode becomes unstable (i.e. there appear the nonvanishing time-oscillations of the conduction current) in the restricted region of applied voltages $2 < U_d < 5 \text{ V}$ while it is stable outside this region [14]. Nevertheless, the dynamic NDR exists above $U_d = 5 \text{ V}$ too. All calculations are performed at the applied voltage $U_d = 5 \text{ V}$ (therefore just inside the stable region of the unloaded diode). Since this point is very close to the unstable region, the main features of the physical processes responsible for the generation must be reflected in the noise characteristics too. Quite different behavior of the noise characteristics is observed under the voltage and current driven operations. Let us consider first the current noise when the constant voltage is applied to the diode.

Fig. 1(a) illustrates the current fluctuations at the short-time scale. The short-time behavior of the current is characterized by spontaneous appearance of super-high frequency oscillations which are spontaneously destroyed in time. The period of oscillations is of about 0.11–0.12 ps that corresponds to the frequency near 8.5 THz. It is the plasma frequency for the n^+ -regions. Thus, the short-time current fluctuations in n^+nn^+ diode are caused by the plasma effects in the n^+ contacts. The long-time behavior of the current fluctuations is shown in Fig. 1(b). To eliminate the plasma oscillations each point $j(t)$ represents the conduction current averaged over the time interval $\Delta t = 0.1 \text{ ps}$ which roughly corresponds to the period of the plasma oscillations. One can observe a spontaneous formation and destruction of the high-frequency oscillations with the period about 4–5 ps. These long-time periodic fluctuations (oscillations) of the current are caused by spontaneous formation of electron accumulation layers and their subsequent drift through the n -region. The layer formation is related to the spatial over-

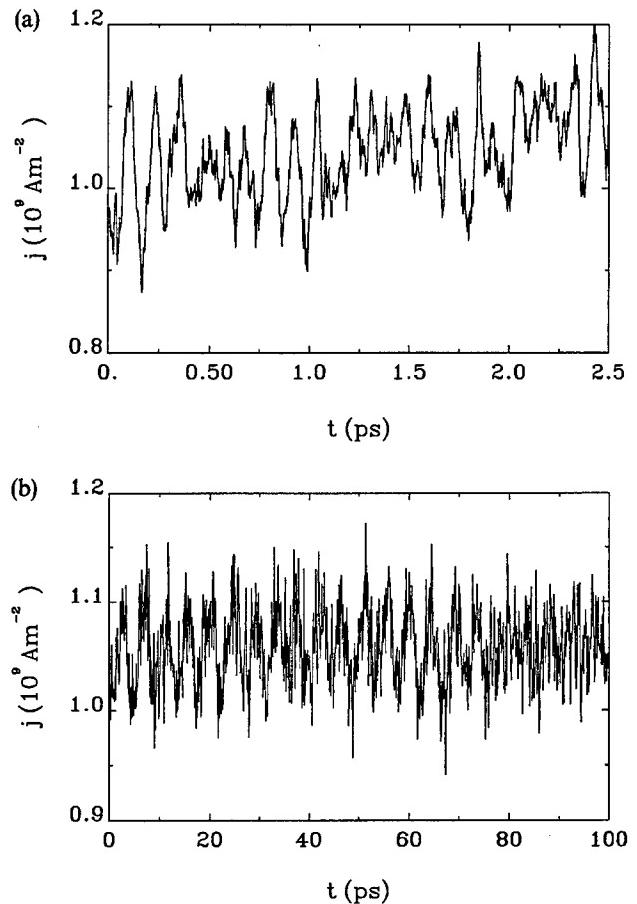


Fig. 1. (a) Short and (b) long time behavior of the current fluctuations obtained by the Monte Carlo simulation under the constant voltage operation for n^+nn^+ InP diode at room temperature with the doping levels of $n = 2 \times 10^{16}$ and $n^+ = 10^{18} \text{ cm}^{-3}$. The cathode, n -region and anode lengths are respectively 0.05, 1.0 and 0.1 μm . The applied voltage $U_d = 5 \text{ V}$.

shoot of the drift velocity due to electron transfer to upper valleys and it usually takes place in the region where the local drift velocity exhibits the maximum negative slope [15]. In the nearmicron n^+nn^+ InP diodes this coincides approximately with the center of the n region. The oscillation frequency is determined by the average transit-time necessary for the layer to be formed, to cross the diode and to disappear in n^+ anode. These oscillations and their relaxation can be explained as follows. Due to the pronounced velocity overshoot in the short InP diodes the average velocity of electrons as function of space firstly sharply increases to its maximum value and then relatively slow decreases reaching a minimum value near the anode contact. Thus, after the maximum the velocity dependence on the spatial coordinate exhibits a negative slope. Spontaneous fluctuations of carrier velocities lead to fluctuation of the local drift velocity and, hence, to fluctuations of the local concentration of carriers. In the diode region characterized by a decreasing profile of the drift velocity, the concentration fluctuations can grow in time, leading to the formation of an accumulation layer. Then the propagation of the layer toward the anode leads to redistribution of the electric field inside the diode and to a smoothing of the velocity overshoot. It is accompanied by a reduction in time of the conduction current in the diode. When the layer leaves the n -region, a redistribution of the internal electric field takes place and the appearance of the pronounced velocity over-

shoot accompanied by the rapid increase in time of the conduction current starts us again. Thus the process of a layer formation, its transport across the diode, and its disappearance at the anode assumes a periodic character which is the basis of microwave generation. However, outside the unstable region (i.e. at $U_d > 5$ V) stochastic character of the carrier velocity fluctuations (both in time and space) destroys the coherence of the successive transit-time oscillations and results in vanishing of their correlation at the sufficiently long times without the feedback, which is realized as a rule by the external resonant circuit (see e.g. [14, 15]).

This short- and long-time behavior of the current fluctuations is responsible for the time behavior of the correlation function of the conduction current fluctuations, $C_j(t)$, at the short- and long-time scales shown in Figs 2(a) and (b), respectively. At the initial part of the time dependence of $C_j(t)$ [see Fig. 2(a)] one observes the damped plasma oscillations which reflect the relaxation of the velocity fluctuations in the n^+ contacts. The coherence in plasma oscillations is fully destroyed per time interval of about 2 ps. It is a usual time of the correlation duration in bulk semiconductors at the room temperatures [10], which is determined primarily by the momentum relaxation time. After the first stage is finished (i.e. at $t > 2$ ps) the correlation function tail demonstrates pronounced oscillations (see Fig. 2(b)) which reflect long-time correlation of the spontaneous transit-time oscillations in the diode.

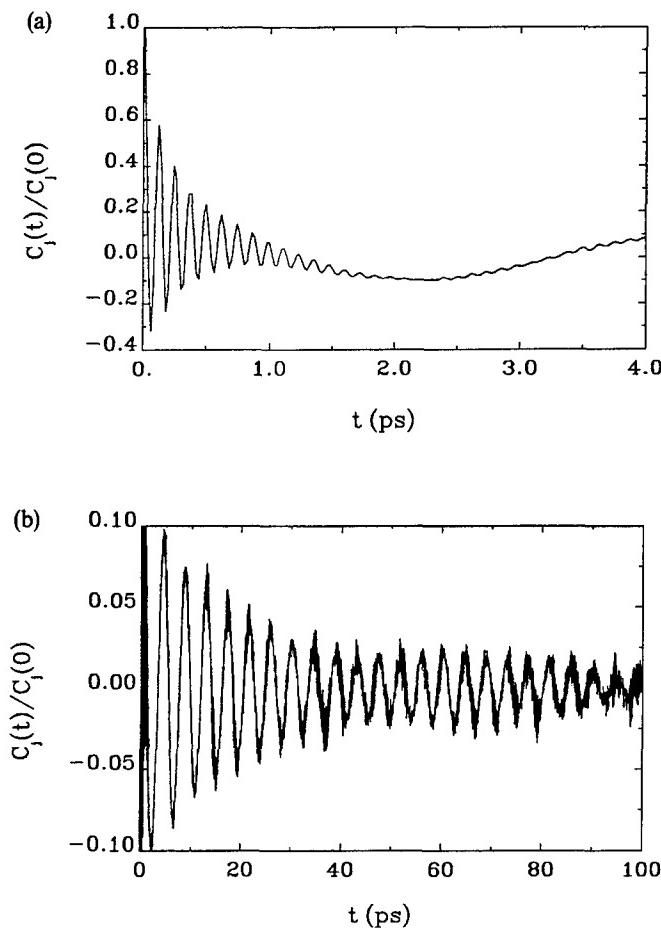


Fig. 2. (a) Short and (b) long time behavior of the correlation function of the current fluctuations calculated for the case of Fig. 1. The $C_j(t)$ is normalized to its initial value at $t = 0$.

Two types of oscillations of $C_j(t)$ in time are responsible for appearance of two peaks in the frequency dependence of the spectral density of the current fluctuations, $S_j(v)$, presented in Fig. 3. The first spike at the frequency $v = 230$ GHz is caused by spontaneous formation of electron accumulation layers and their subsequent drift through the n -region. The second one, i.e. the resonant peak of $S_j(v)$ near the frequency $v = 8$ THz corresponds to the plasma oscillations in the n^+ regions. Comparing the first peak position with the whole amplification range $v = 90$ – 230 GHz at $U_d = 5$ V determined for the same diode from the small-signal analysis performed by the hydrodynamic approach [14, 15] one can conclude that the spontaneous transit-time oscillations correspond to the maximum generation frequency and, hence, to minimum transit-time necessary for a single accumulation layer to be formed, to cross the n region, and to vanish at the anode. This frequency is in good agreement with the maximum frequency of microwave power generation observed experimentally in similar diodes [12].

Let us consider now the voltage noise characteristics of the same diode calculated at the constant total current $j = 1.03 \times 10^9$ A/m², corresponding to the average applied voltage $U_d = 5$ V. Let us recall that the constant current operation implies existence of very high external resistance connected in series with the diode. Therefore, under the conditions close to the current driven operation all conduction current and voltage oscillations must be effectively damped in time due to influence of this resistance. It is in a good agreement with the short- and long-time behavior of the voltage fluctuations presented in Figs 4(a) and (b), respectively. One can observe a rather small (less than 1 percent) plasma oscillations of the voltage drop between the diode terminals, which are superimposed on a long-time variation of U_d . The long-time fluctuations of U_d are of the same order in magnitude as the current fluctuations considered above [cf. Fig. 2(b)]. However, any periodicity in the fluctuations is absent. It is evidenced by Fig. 5 where the correlation function of the voltage fluctuations is shown as a function of time. The $C_U(t)$ decreases monotonically (but not exponentially) with time and practically goes to zero per time $t = 4$ – 5 ps, that coincides with the average transit-time of a single accumulation layer. Thus the average transit-time can be considered as a correlation duration of internal fluctuations inside the diode under the constant current oper-

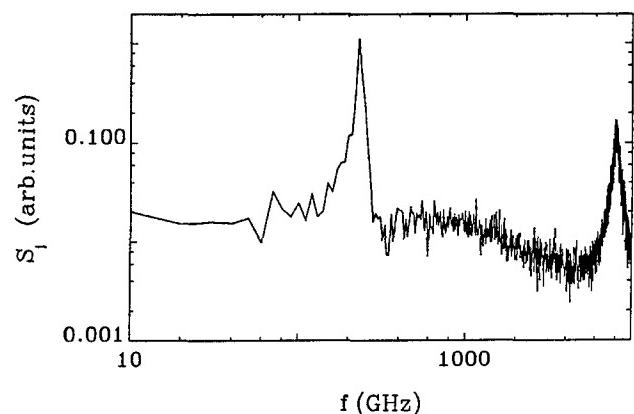


Fig. 3. Frequency dependence of the spectral density of the current fluctuations calculated from the correlation functions presented in Fig. 2. $U_d = 3$ V.

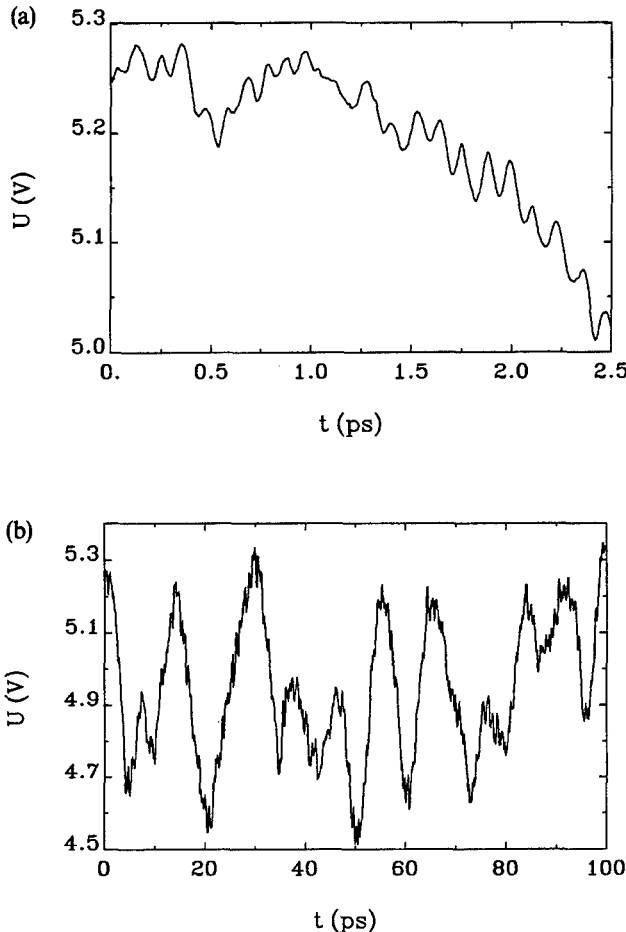


Fig. 4. (a) Short and (b) long time behavior of the voltage fluctuations obtained by the Monte Carlo simulation under the constant current operation for n^+nn^+ InP diode at the average applied voltage $U_d = 5$ V. The diode parameters are the same as Fig. 1.

ation. The monotonous decrease of the correlation function results in the Lorentzian shape of the frequency dependence of the spectral density of the voltage fluctuations shown in Fig. 6 by solid line. The almost linearly decrease of the correlation function at small-time scale with sufficiently long characteristic time (of about 10 ps) cuts the high-frequency part of the spectrum. Such a shape of the voltage noise spectrum implies that the long-time fluctuations of the voltage with duration of about 10 ps and more are mostly responsible for the noise. It is in a good agreement with the time behavior of the voltage fluctuations presented in Fig. 4(b). For comparison Fig. 6 also presents the first peak of the spectral density of the current fluctuations calculated under the constant voltage operation (both spectra are normalized to their maximum values). The spectrum of the current fluctuations outside the resonant region remains nearly flat up to the frequencies of about 1 THz which is roughly corresponds to duration of the correlation on the microscopic level caused by the combined action of the momentum and energy relaxation times [10, 16]. Indeed, the time dependence of the current fluctuations is much more noisy than that of the voltage fluctuations on the time scale less than 10 ps [cf. Figs 1(b) and 4(b)]. Moreover, the transit-time resonant peak appears when the voltage noise practically disappears. Thus, for the same diode the voltage and current noise take place in different frequency ranges and has different character inspite of the fact that they both

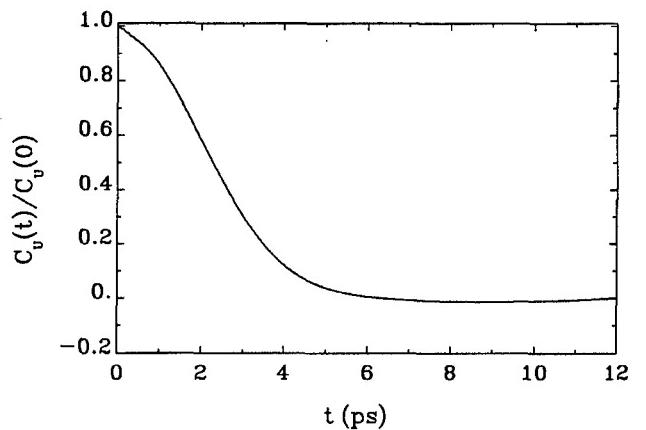


Fig. 5. Time behavior of the correlation function of the voltage fluctuations at the diode terminals calculated for the case of Fig. 4. The $C_v(t)$ is normalized to its initial value at $t = 0$.

have the same microscopic cause related to the velocity fluctuations of single particles.

To explain the main peculiarities of such a behavior it is necessary to take into account that the diode noise which is induced in the external circuit is determined not only by distribution of the internal microscopic sources but also by a capability of the diode to amplify or to damp small perturbations appearing inside the diode. For the voltage and current driven operations these capabilities are described in a macroscopic level by the small-signal admittance and impedance of the diode, $Y(v)$ and $Z(v)$, respectively, which are mutually connected by a relation $Y(v)Z(v) = 1$. As it follows from eq. (4) the square modulus of these quantities give interconnection between the frequency dependencies of the current and voltage noise. Therefore, Fig. 7 presents the frequency dependence of the absolute-values squares of the small-signal impedance and admittance of the diode calculated by the HD approach. The $|Z(v)|^2$ decreases monotonically with the frequency v while the spectrum of $|Y(v)|^2$ exhibits the sharp spike which is a consequence of the described above resonant behavior of the diode under the voltage driven operation. The resonant frequency corresponds to the upper boundary of the first generation band [14], that is to the minimum time which is required by the accumulation layer to transit across the diode. By comparing Figs 6 and 7 we conclude that the main features of the

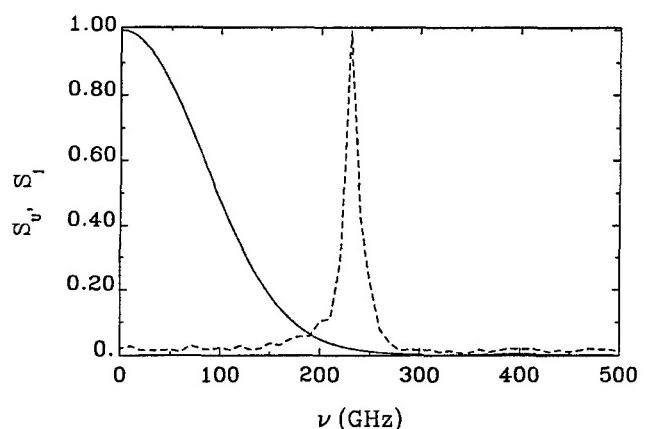


Fig. 6. Frequency dependence of the spectral density of the voltage and current fluctuations (solid and dashed lines, respectively).

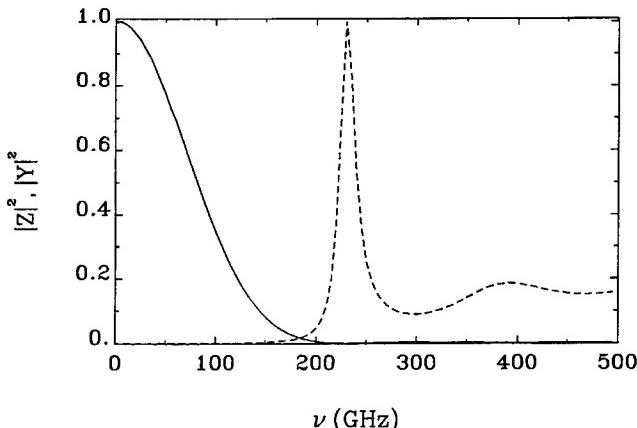


Fig. 7. Frequency dependence of the square modulus of the small-signal impedance and admittance calculated by the hydrodynamic approach (solid and dashed lines, respectively). The diode parameters are the same of Fig. 1.

spectral characteristics of the equivalent sources of the current and voltage noise are caused by the frequency dependence of $|Y(v)|^2$ and $|Z(v)|^2$, respectively. The resonant behavior of the small-signal admittance is responsible for the peak of the spectral density of the current fluctuations near the transit-time frequency and, in turn, the monotonous decrease of the spectral density of the voltage fluctuations at increasing frequency is caused by a similar behavior of $|Z(v)|^2$.

4. Conclusions

By performing a theoretical analysis of the noise spectral density of submicron n^+nn^+ InP diodes, we have proven that the calculation of the noise spectra provides useful information about the physical processes responsible for the device performance. Under the constant voltage operation the microscopic fluctuations of the single particle velocities in the n and n^+ regions results in macroscopic oscillations of the correlation function of the current fluctuations at the both long- and short-time scales leading to the resonant peaks in the noise spectra at the transit-time and plasma frequencies, respectively. Outside the transit-time resonance

region the current noise is nearly white up to frequencies of about 1 THz, which corresponds to correlation duration caused by the momentum and energy relaxation times. The transit-time resonance is caused by quasi-periodic process one cycle of which includes a formation of the electron accumulation layer near the n region center, its travelling across the diode and disappearance at the anode contact. Under the constant current operation the voltage noise spectrum has the usual Lorentzian shape which starts at low frequencies and practically vanish at the transit-time frequency.

Acknowledgement

The authors would like to express their sincere gratitude to the Baltic Scholarship Scheme of the Nordic Council of Ministers.

References

1. Price, P. J., "Fluctuation Phenomena in Solids" (Edited by R. E. Burgess) (Academic Press, New York 1965), p. 355.
2. Reggiani, L. (Ed.), "Hot Electron Transport in Semiconductors" (Springer, Berlin 1985), Topics of Applied Physics, vol. 58.
3. Van der Ziel, A., "Noise in Solid State Devices and Circuits" (Wiley, New York 1986).
4. Cappy, A., IEEE Trans. Microwave Theory Tech., **36**, 1 (1988).
5. Nougier, J. P., in: "III-V Microelectronics" (Edited by J. P. Nougier) (Elsevier Science Publishers B.V. 1991), p. 183.
6. Filicori, F., Ghione, G. and Naldi, C. U., IEEE Trans. Microwave Theory Tech., **40**, 1333 (1992).
7. Zimmermann, J. and Constant, E., Solid-State Electron., **23**, 915 (1980).
8. Lugli, P., Reggiani, L. and Niez, J. J., Phys. Rev. **B40**, 12382 (1989).
9. Reggiani, L., Kuhn, T. and Varani, L., Appl. Phys. A, **A54**, 411 (1992).
10. Starikov, E. V. and Shiktorov, P.N., Liet. Fiz. Rink. **32**, 471 (1992).
11. Mitin, V., Gružinskis, V., Starikov, E. and Shiktorov, P., J. Appl. Phys., **75**, 935 (1994).
12. Rydberg, A., Electr. Dev. Lett. **11**, 439 (1990).
13. Kamona, R., Eisele, H. and Haddad, G. I., Solid State Electr. **36**, 1547 (1993).
14. Gružinskis, V. et al., "Simulation of Semiconductor Devices and Processes" (Edited by S. Selberherr, H. Stippel and E. Strasser) (Springer-Verlag, Wien 1993), vol. 5, p. 333.
15. Gružinskis, V., Starikov, E., Shiktorov, P., Reggiani, L. and Varani, L., Phys. Rev. **B49**, (1994), in press.
16. Gružinskis, V., Starikov, E. and Shiktorov, P., Solid State Electr. **36**, 1067 (1993).

Effect of Partial Ionization and the Characteristics of Lateral Power Diamond MESFETs

K. J. Grahn, P. Kuivalainen and S. Eränen

Technical Research Centre of Finland, Electronics, IC Design, P.O. Box 11012, FIN-02044 VTT, Finland*

Received May 4, 1994; accepted in revised form July 8, 1994

Abstract

Power device structures, which could be implemented by using diamond technology of today, have been analyzed with the aid of a two-dimensional numerical simulator code for semiconductor devices. It has been found that partial ionization of deep acceptor states substantially degrades the electrical performance of 1000 V lateral diamond power MESFETs. No advantage over corresponding silicon devices can be obtained. If the partial ionization phenomenon can be eliminated, the on-resistance of the diamond MESFET is almost two orders of magnitude smaller than in vertical silicon power MOSFETs having the same breakdown voltage.

1. Introduction

Recent progress in silicon power MOSFETs has brought the on-resistance down to within a factor of 2 of the theoretical minimum value $0.13 \text{ m}\Omega\text{cm}^2$ for 55 V devices [1]. It is therefore, necessary to consider other semiconductor materials if further improvements in device performance are to be realized in the future.

Recently it has been recognized that semiconducting diamond is a very attractive material for high-power devices due to its high breakdown field (10^7 V cm^{-1}), high electron and hole mobilities (2000 and $1800 \text{ cm}^2/\text{Vs}$, respectively), and the highest thermal conductivity ($20 \text{ W cm}^{-1} \text{ K}^{-1}$) of any solid at room temperature [2]. However, the performance of diamond devices made so far has been poor [3–6]. This is partly due to difficulties in single crystal growth, material etching, selective doping and metal contact formation. Also dimensions and layouts of the diamond devices have been far from optimized ones.

Wide bandgap semiconductors typically can be doped either n-type or p-type, but not both [7]. Doped layers have so far been produced with limited success [8]. A major difference between the atomic concentration and the carrier concentration for each dopant has been found. One possible explanation for this phenomenon is that defects compensate dopants. Partial ionization of deep acceptor (donor) states will therefore affect the electrical performance of the material.

In this paper we show by using a semiconductor device simulation code that partial ionization of deep acceptor states deteriorates severely the otherwise excellent electrical performance of lateral power diamond MESFETs. The results are also compared to those of a corresponding vertical power silicon MOSFET.

2. Simulation model

When investigating the potential of diamond in power device applications by using 2D numerical simulation we rely on the fact that the theory of physical properties of semiconductors and of the relation of these properties to the performance of devices is well understood and this theory can be applied to any semiconductor material. This actually provides a basis of the comparison and ranking of devices made of different semiconductor materials.

In the simulations of the electrical characteristics of diamond power transistors we use the two-dimensional simulator code SCORPIO [9], which was originally developed for silicon devices. SCORPIO solves the Poisson equation and the current continuity equations in the drift diffusion approximation for electrons and holes. The Governing equations have been spatially discretized by use of the Box Integration Method (BIM) [10]. The BIM is a generalization of the finite difference method and it requires an appropriate partitioning of the device domain into boxes. The device is built up by rectangles and/or acute triangles defined by the user and the mesh is automatically generated and refined on the basis of these elements. The Poisson equation is integrated over the box and the divergence of the electric field is approximated with a weighted sum of the components of the field to all neighboring grid points. For the continuity equations the well-known Scharfetter-Gummel discretization scheme [11] is applied. The integral of the divergence of the current density is approximated in the same way as that of the electric field. The electric field and the mobility are calculated in a virtual grid point between two ordinary grid points.

The analytical form of the basic transport equations is independent of the semiconductor material. Differences in the simulated results are due to physical model functions, which describe the semiconductor material dependent parameters such as mobility, diffusivity, and generation and recombination of the charge carriers. On the other hand, the model functions depend on the charge carrier density, current density and electric field, which makes the basic transport equations highly non-linear.

The relatively large activation energies of the electrically active impurities (donors and acceptors) in diamond have raised an important question about the complete ionization of the impurities. In power semiconductor devices the performance is often predominantly determined by the lowly doped regions, which are required for the voltage capability

* email: kaj.grahn@vtt.fi

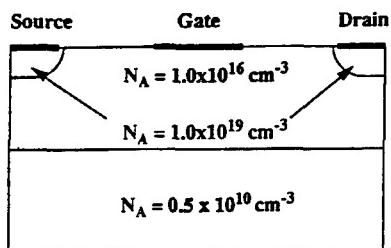


Fig. 1. Cross-section of a lateral diamond MESFET used in the simulations. The gate length is $2\mu\text{m}$ and both the drain-gate and source-gate distances are $3\mu\text{m}$. The thickness of the active diamond layer having the doping $N_A = 10^{16}\text{ cm}^{-3}$, is $1.5\mu\text{m}$.

of the devices. By writing $n = N_D^+$ we can derive an expression for the density of the ionized donor levels N_D^+ [12]

$$N_D^+ = \frac{2N_D}{1 + 4\sqrt{\frac{1}{16} + \frac{1}{2} \frac{N_D}{N_C} \exp\left(\frac{E_a}{k_B T}\right)}} \quad (1)$$

Here N_D is the density of the donor atoms, N_C the effective density of states in the conduction band and E_a the activation energy of the donor levels. A similar equation can be written for ionized acceptors but double degeneracy of the impurity level has to be taken into account. For some typical values in diamond, $N_D = 10^{15}\text{ cm}^{-3}$, $N_C = 10^{19}\text{ cm}^{-3}$, and $E_a = 0.2\text{ eV}$, we have $N_D^+ = 0.74N_D$ at $T = 300\text{ K}$. Thus the ionization is almost complete even at room temperature. However, experimental results [8] in p-type diamond thin films show that only a small fraction of boron acceptors is ionized in CVD. In the simulations of the lateral p-MESFETs we have taken this effect into account.

Our model functions for the mobility of the electrons and holes include carrier velocity saturation at high fields as well as the influence of temperature, doping and carrier densities, and electric field [9, 13]. In the generation and recombination model we include the Shockley-Read-Hall, Auger and impact ionization mechanisms [13]. In order to consider the thermionic nature of the current at the metal-semiconductor interface, we write the boundary conditions

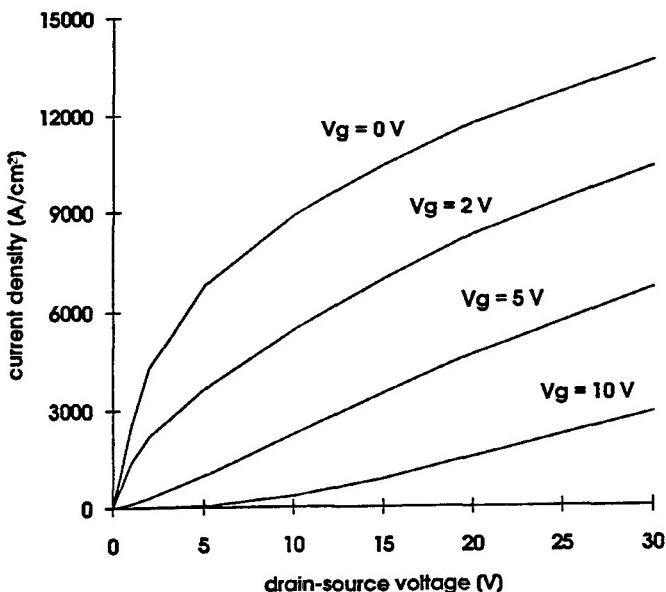


Fig. 2. $I-V$ characteristics of the p-type diamond MESFET shown in Fig. 1.

for the hole densities at the interface as [14]

$$J_p = A_{R,p}^{\text{eff}} T^2 \frac{(p - p_s)}{N_v} \quad (2)$$

where p_s is the equilibrium surface concentration of holes, J_p is the hole current density and N_v is the effective density of states in the valence band. The surface concentration of holes is calculated from the surface potential. Further, $A_{R,p}^{\text{eff}} = 1.4 * A_R$ is the effective Richardson constant of holes.

3. Simulations on lateral power diamond MESFET

Vertical power device structures have higher power handling capability than lateral structures. However, in the case of diamond there are severe problems related to the fabrication of n-type vertical devices. All the diamond transistors made so far have been lateral p-type devices. Therefore, we have analyzed the lateral MESFET structure shown in Fig. 1, and optimized it as a power transistor. The MESFET structure was chosen since a Schottky junction is a more realistic gate structure in diamond than a pn-junction in a JFET. The p-type device was also chosen because of difficulties in growing doped n-type diamond epitaxial layers. So, the chosen device structure could be implemented by using existing diamond and semiconductor technologies.

Dimensions of the analyzed MESFETs have been chosen in order to compare transistor characteristics with vertical silicon devices in power applications. The minimum value for the drain-gate distance L_{dg} , in the MESFET structure shown in Fig. 1 is determined by the wanted breakdown voltage V_{br} , and also by the space charge limited current (SCLC) [15]. We have found SCLC to become dominant in very small diamond device structures. On the other hand, large values of L_{dg} increase the on state resistance R_{on} , and therefore an optimized value for L_{dg} must be found. When the simulated V_{br} is just above 1000 V , the minimum value for L_{dg} is $3\mu\text{m}$. In this case R_{on} is as small as $7\text{ m}\Omega\text{cm}^2$. This value is about two orders of magnitude smaller than the theoretical minimum value $266\text{ m}\Omega\text{cm}^2$ for vertical 1000 V -silicon power MOSFETs [16]. This difference is due to the fact that the high breakdown field of diamond allows a much higher doping density.

Figure 2 shows the whole $I-V$ characteristics of the optimized diamond MESFET. The gate control is good, i.e., the transistor can be switched to an off-state with a moderate change in the gate voltage. Transconductance in this p-type diamond transistor is 8.6 mS/mm at $V_{ds} = -10\text{ V}$ and $V_{gs} = 1\text{ V}$.

The simulation results indicate that even lateral diamond devices would be superior to the most advanced silicon power devices in the same V_{br} rating. Notice that we have not compared thermal properties of the devices. These considerations would even strengthen the above conclusions, since thermal conductivity of diamond is an order of magnitude larger than in silicon.

4. Effect of partial ionization

In the simulations above it was assumed that all the dopant atoms will be ionized. In the case of rather large ionization energies of dopants in diamond, almost all donors or accep-

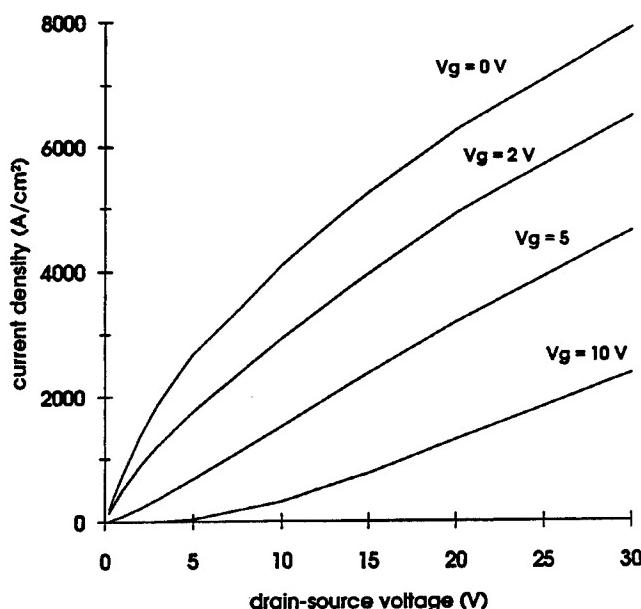


Fig. 3. $I-V$ characteristics of the modified diamond MESFET.

tors seem to be ionized in the lightly doped regions of the device according to the simple model discussed in Section 2. However, experimental results [8] for p-type diamond thin films indicate, that only a very small fraction (10^{-3}) of impurities is ionized. It is not yet certain that this drawback can be eliminated in the future as the diamond technology develops. Therefore we have performed simulations on lateral diamond p-MESFET structures by taking this effect into account.

In the case of power transistors the partial ionization means that in order to get a low on-state resistance with a carrier concentration of 10^{16} cm^{-3} , the doping concentration should be as high as 10^{19} cm^{-3} . On the other hand, the heavy doping decreases the breakdown voltage and results in a very low mobility, which in turn increases the on-state resistance. Therefore it is not obvious in advance, that diamond would remain a superior material for power transistors, when the partial ionization is taken into account.

Since only a fraction of impurities is ionized, the high dopant concentrations should not decrease V_{br} as dramatically as in the normal completely ionized case. On the other hand, the avalanche breakdown should occur much easier for these unionized impurities with ionization energies 0.2–0.3 eV than in the ordinary avalanche process related to the generation of electron-hole pairs across the 5.5 eV band gap. We have simulated the worst case, i.e., by assuming that due to doping, V_{br} decreases as much as in the ordinary case with fully ionized impurities.

The simulated lateral diamond MESFET is the same as in Fig. 1 but the doping density of the active layer is now chosen to be $N_A = 10^{19}\text{ cm}^{-3}$. We assume that only 1/1000 of the acceptors in the active layer are ionized. However, the new acceptor doping degrades the mobility. The pinch-off voltage will not change because only ionized impurities

affect its values. For high gate voltages we can expect ionization of impurities beneath the gate due to the high electric field. This affects the driving capability and the on-resistance of the device. In order to minimize this effect we have calculated the on-resistance at a gate-source voltage of $V_{GS} = 0\text{ V}$.

Figure 3 shows the $I-V$ characteristics of the modified diamond MESFET. The current values are about 2/3 of those calculated for the original lateral diamond MESFET. The calculated R_{on} is $24\text{ m}\Omega\text{ cm}^2$ and the calculated V_{br} is 110 V for a doping density of $N_A = 10^{19}\text{ cm}^{-3}$. For a corresponding power Si MOSFET (100 V) with a gate length of $L_g = 0.3\text{ }\mu\text{m}$ the on-resistance is $4\text{ m}\Omega\text{ cm}^2$ [16].

5. Conclusions

Computer simulations show that partial ionization of deep acceptor states substantially degrades the performance of lateral diamond power MESFETs. The results indicate that we have no advantage over corresponding silicon devices. We also show that if the partial ionization drawback can be eliminated, the simulated electrical performance of lateral diamond power MESFETs exceeds by orders of magnitude those of silicon devices having the same breakdown voltage. Additional advances of diamond devices are high thermal conductivity and high temperature operation. The simulated structures can be fabricated by using the existing diamond device technology. Some improvements in the performance can still be expected when changing from silicon to diamond, but this necessitates a significant progress in the fabrication technologies of diamond devices.

References

1. Shenai, K., IEEE Electron Device Lett. **12**, 108 (1991).
2. Shiomi, K., Scott, R. S. and Baliga, B. J., IEEE Trans. Electron Dev. **36**, 1811 (1989).
3. Shiomi, H., Nishibayashi, Y. and Fujimori, N., Jap. J. Appl. Phys. **28**, L2153 (1989).
4. Gildenblat, G. S. H., Grot, S. A., Hatfield, C. W. and Badzian, A. R., IEEE Electron Dev. Lett. **12**, 37 (1991).
5. Tsai, W. et al., IEEE Electron Dev. Lett. **12**, 157 (1991).
6. Zeisse, C. R. et al., Electron Dev. Lett. **12**, 602 (1991).
7. Laks, D. B., Van de Walle, C. G., Neumark, G. F. and Pantelides S. T., Phys. Rev. Lett. **66**, 648 (1991).
8. Davis, R. F. et al., Mat. Sci & Eng. **B1**, 77 (1988).
9. Grahn, K. J., Acta Polytechnica Scandinavica (Electrical Engineering Series) **76**, 36 (1993).
10. Hitschfeld N., Series in Microelectronics **21**, 5 (1993).
11. Scharfetter, D. L. and Gummel, H. K., IEEE Trans. Electron Dev. **16**, 64 (1969).
12. Seeger, K., "Semiconductor Physics" (Springer Verlag, Wien/New York 1973), p. 45.
13. Engl, W. L., Dirks, H. K. and Meinerzhagen B., Proc. IEEE **71**, 10 (1983).
14. Gaur, S. P. et al., IBM J. Res. Dev. **29**, 242 (1985).
15. Lambert, M. A. and Mark, P., "Current Injection in Solids" (Academic Press, New York 1970), p. 27.
16. Baliga, B. J., "Modern Power Devices" (John Wiley & Sons, New York 1987), p. 296.

Quantum Corrections to the Threshold Voltage of Short Channel MOSFETs

P. Kuivalainen

Technical Research Centre of Finland, Electronics, IC Group, P.O. Box 11012, FIN-02044 VTT, Finland

Received May 2, 1994; accepted in revised form June 28, 1994

Abstract

An analytical model is developed which accounts the quantum mechanical corrections to the threshold voltage V_T MOSFETs having their channel lengths in deep submicron region. The model is based on a variational solution of the Schrödinger equation for electrons in an inversion layer of a MOSFET, and it takes into account effects of the quantized electrons, the electron charge distribution and bandgap narrowing on V_T . Without any fitting parameters the model explains well the measured quantum shifts of V_T .

1. Introduction

It is well known [1], that in the case of strong inversion, electron motion in the channel of a MOSFET is quantized in the direction perpendicular to the Si-SiO₂ interface. In the modelling of the electrical behaviour of the MOSFETs this quantization usually can be neglected at room temperature. However, in the modern MOSFETs having their channel lengths in the deep submicron region, the channel doping concentrations have been increased in order to suppress punch-through currents. This, in turn, enhances the quantum effects in the channel region by increasing the splitting between the quantized energy levels of the charge carriers. The quantization increases the threshold voltage of the MOSFET. This is due to two facts: In the quantized case the first allowed energy level E_0 does not coincide with the bottom of the conduction band E_C . Therefore, due to the energy difference $E_0 - E_C$, an additional band bending is needed for the onset of the inversion in the MOSFET channel. On the other hand, when the charge density in the channel is considered, there is a difference between the classical and quantum solutions. The average distance of the charge density from the Si-SiO₂ interface is larger in the quantized case, thereby decreasing the effective oxide capacitance and further increasing the threshold voltage V_T . In deep submicron devices the threshold voltage must be designed to be below 0.6 V, which also enhances the importance of the quantum mechanical corrections to V_T in the modelling of these devices.

The first analytical model for the quantum corrections of V_T was published recently by van Dort *et al.* [2, 3]. They proposed a simple model for the extra band bending, $\Delta\Psi \sim E_0 - E_C = \beta(N_A)^{1/3}$, where N_A is the acceptor doping concentration and β a fitting parameter. The model explained well the measured changes in V_T , although in some cases [2]

unphysically large values for the fitting parameter β were needed.

In this paper we compare two analytical models for the quantum mechanical corrections to V_T based on two first order analytical solutions of the Schrödinger equation ([1]: A variational solution and the triangular potential approximation. The calculated corrections to V_T are compared to the experimental results without any fitting parameters.

2. Theory

The classical expression [4] for the threshold voltage of a n-channel MOSFET is given by

$$V_T^{CL} = F_{FB} + \Psi_s + \frac{\sqrt{2\epsilon_{Si} q N_A \Psi_s}}{C_{ox}} \quad (1)$$

where F_{FB} is the flat band voltage including work function difference between silicon and gate material, and interface and oxide charges. Ψ_s is the surface potential in the Si-SiO₂ interface, N_A the substrate doping concentration, and $C_{ox} = \epsilon_{SiO_2}/t_{ox}$ is the oxide capacitance per unit area, where ϵ_{SiO_2} is the permittivity of SiO₂ and t_{ox} the oxide thickness.

In the classical solution of the Poisson equation the electron density in the channel has a maximum at the silicon surface, whereas in the quantum mechanical solution the maximum is shifted from the surface [1]. This causes a change in the effective capacitance [2, 3]

$$C_{ox}^{QM} = \frac{\epsilon_{ox}}{t_{ox} + \frac{\epsilon_{ox}}{\epsilon_{Si}} \Delta z} \quad (2)$$

where, $\Delta z = \langle z \rangle_{QM} - \langle z \rangle_{CL}$ is the difference between the quantum mechanically and classically calculated average distances to the interface. The silicon bandgap is effectively increased due to energy difference between the bottom of the conduction band E_C and the first allowed energy level E_0 in the quantized channel. The total band bending due to the quantum corrections is now given by

$$\Delta\Psi_s = (E_0 - E_C - \Delta E_g)/q + F_s \Delta z \quad (3)$$

where the ordinary bandgap narrowing [5], $\Delta E_g \approx 18 \text{ meV} \cdot \log(N_A/10^{17})$, due to heavy doping partially compensates the bandgap widening $E_0 - E_C$. F_s is the surface electric

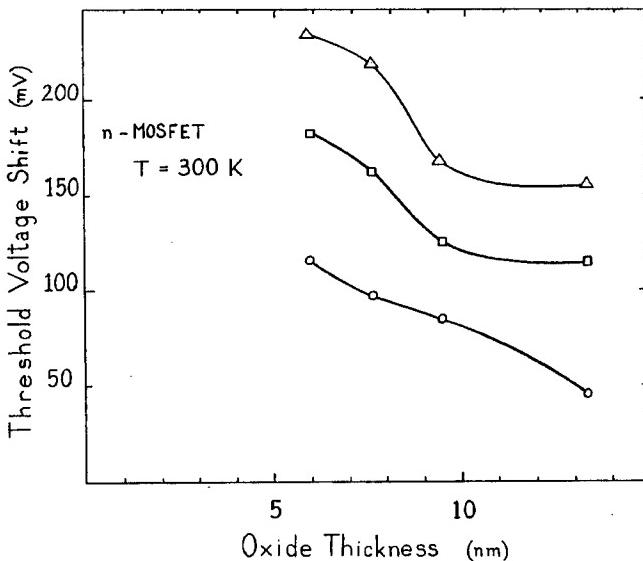


Fig. 1. Quantum mechanical shifts of the threshold voltage vs. oxide thickness in deep submicron MOSFETs. The experimental results (circles) are from [3], and the theoretical results have been calculated by using the variational approach (squares) and triangular potential approximation (triangles).

field given by

$$F_s = \frac{q}{\epsilon_{Si}} (N_{depl} + n_{inv}) \quad (4)$$

where N_{depl} is the depletion charge density, and n_{inv} the electron density in the inversion layer. By using (4), the classical value for $\langle z \rangle_{CL} = kT/qF_s$, can be calculated [1]. By using (2) and (3) the quantum mechanically corrected threshold voltage can be written as

$$V_T^Q = V_{FB} + \Psi_s^Q + \frac{\sqrt{2\epsilon_{Si} q N_A \Psi_s^Q}}{C_{ox}^Q} \quad (5)$$

where $\Psi_s^Q = \Psi_s + \Delta\Psi_s$.

So, in order to estimate V_T^Q , we have to calculate $E_0 - E_C$ and Δz in (3), which can be obtained from the solution of the Schrödinger equation [1]

$$\frac{\hbar^2}{8\pi^2 m^*} \frac{d^2}{dz^2} \Phi(z) + (E_i - V(z))\Phi(z) = 0 \quad (6)$$

where $V(z)$ is the electrostatic potential in the channel and it is obtainable from Poisson equation

$$\frac{d^2}{dz^2} V(z) = \frac{q N_A}{\epsilon_{Si}} \quad (7)$$

where N_A is the acceptor concentration in the channel.

A self-consistent solution to the coupled Poisson and Schrödinger equations is possible only by numerical means. However, the limit, in which only one subband has carriers in it, can be solved by using a variational method [1]. The wave function $\Phi(z)$ can be approximated by a trial function

$$\Phi(z) = \left(\frac{\alpha^3}{2}\right)^{1/2} z e^{-\alpha z/2} \quad (8)$$

If we neglect charge density n_{inv} in the inversion layer, the energy of the lowest subband is given by

$$E_0 - E_C = \left(\frac{3}{2}\right)^{5/3} \left(\frac{q^2 h}{2\pi\epsilon_{Si}}\right)^{2/3} \frac{N_{depl}^{2/3}}{m^{*1/3}} \quad (9)$$

The average value of z for the variational wave function (8) is given by $\langle z \rangle_{QM} = 3/\alpha$, where α is the value that minimizes the total energy

$$\alpha = \left(\frac{48\pi^2 m^* q^2 N_{depl}}{\epsilon_{Si} h^2}\right)^{1/3} \quad (10)$$

An alternative analytical approach for solving (6) is based on the approximation where the potential energy $V(z)$ is described by a triangular potential [1], $V(z) = qF_s z$ for $z > 0$, and elsewhere $V(z)$ is assumed to be infinite. This leads to the Airy equation with the following solution for the lowest subband

$$E_0 - E_C = \left(\frac{h^2}{8\pi^2 m^*}\right)^{1/3} \left(\frac{9\pi q F_s}{8}\right)^{2/3} \quad (11)$$

The average value of z is now given by $\langle z \rangle_{QM} = 2E_0/3qF_s$. The threshold voltage shift $\Delta V_T = V_T^Q - V_T^{CL}$ can now be calculated in two simple cases, eqs (9) and (11). The ratio between the solutions (11) and (9) is 1.180046, i.e. the variational approach leads to a lower energy.

3. Results

Dort *et al.* [2, 3] have estimated experimentally the quantum shifts of V_T by comparing the measured V_T 's to those obtained from simulations when using a classical 2-D device simulator. A special care was taken in order to eliminate other possible explanations for the observed large shifts in V_T .

Figure 1 shows the calculated and measured [3] results for the shift in the threshold voltage as a function of the oxide thickness in deep submicron MOSFETs having their channel lengths in the range of 0.17 to 0.5 μm. Our calculated results have been obtained both for the variational, solution (9) and the triangular potential approximation (11). The former gives lower values for ΔV_T , as it should since the variational approach gives the lowest energy and thereby a smaller shift of V_T . The dependence of ΔV_T on the oxide thickness (and the doping concentration N_A) is well described by the theory, but simple models (9) and (11) give larger shifts than the experimentally observed ones. However, in another series of experiments Dort *et al.* [2] have observed larger shifts, even in the range of 100–250 mV. Therefore we may state that the simple models give the right order of magnitude for ΔV_T . The theoretical results for ΔV_T presented by Dort *et al.* [2, 3] are clearly smaller (in the range 50–110 mV for $t_{ox} = 5.9$ –13.3 nm) than our results shown in Fig. 1. The reason for this difference is, e.g. the approximation they made for the threshold voltage expression (5).

4. Conclusions

Simple analytical models for the quantum mechanical corrections to the threshold voltage V_T in deep submicron MOSFETs, can easily be derived. The models explain the correct order of magnitude for the measured shift ΔV_T and the right dependences of ΔV_T on doping concentration and

oxide thickness. The results show that in deep submicron devices the quantum mechanical corrections are large and they certainly must be taken into account in accurate modelling.

References

1. Stern, F., CRC Critical Reviews., in: "Solid State Sciences," 499, May 1974.
2. van Dort, M. J. and Woerlee, P. H., in: "Simulation of Semiconductor Devices and Processes" (Edited by W. Fichtner, and D. Aemmer) (Hartung-Gorre Verlag 1991), Vol. 4, p. 451.
3. van Dort, M. J., Woerlee, P. H., Walker, A. J., Juffermans, C. A. H. and Lifka, H., IEEE Proceedings of the International Electron Devices Meeting 1991, p. 495.
4. Muller, R. S. and Kamins, T. I., "Device Electronics for Integrated Circuits" (John Wiley & Sons, New York 1977), p. 323.
5. Slotboom, J. W. and de Graaff, H. C., Solid-State Electronics, **19**, 857 (1976).

Physical Modelling of Vertical DMOS Power Transistors for Circuit Simulation

M. Andersson and P. Kuivalainen

Technical Research Centre of Finland, Electronics, IC Group, P.O. Box 11012, FIN-02044 VTT, Finland

Received May 2, 1994; accepted June 14, 1994

Abstract

A physical model for vertical DMOS power transistors is presented. The model takes into account various short channel effects in the DMOS channel region and the velocity saturation and the exact device geometry in the drift region. The model, aimed at computer aided design of power integrated circuits, has been implemented in the APLAC circuit simulator. A good agreement between the measured and simulated results for vertical DMOSTs is demonstrated.

1. Introduction

In the power integrated circuits (PIC) low voltage logic and MOS-gated power devices have been integrated on the same silicon chip. The computer aided design of the PICs is typically a tedious analog task, where accurate device models are required in order to guarantee reliable simulations. Standard SPICE transistor models [1] as such cannot describe the unique features of the new power devices, such as quasisaturation at high bias levels. In physical device models these new features can be taken into account, although the implementation of the model may require a great deal of work. Recently Kim and Fossum (K-F) [2] published a physical model for vertical power DMOSTs. In the present paper we propose an alternative model, which differs from the K-F model as follows: In the channel region the effect of the acceptor doping gradient on the threshold voltage is treated in a more general case, and the velocity saturation of charge carriers is described in accordance with the measured results for short channel MOSFETs. In the accumulation region the space charge limited current is neglected, but instead, the voltage drop is estimated directly by integrating the electric field over the region. This simplifies the model and decreases strongly the computing time.

2. Model

A new feature in the power MOSFETs, which is not present in the standard low voltage devices, is the so-called quasisaturation, i.e., the loss of gate control of the drain current at very high current levels and voltages. In order to describe quasisaturation, the velocity saturation of the charge carriers must be taken into account in the modelling of the drain I_D

$$I_D = \frac{A_y q n \mu_n E_y}{1 + \frac{E_y}{E_C}} \quad (1)$$

where n is the electron concentration, μ_n the low-field electron mobility, $E_C = v_s/\mu_n$, and v_s the electron saturation

velocity. Figure 1 shows the cross-section of a vertical DMOST, where the drift region has been divided into three parts: an accumulation region (A), a drift region with a varying cross-section area A_y (B), and a drift region with a constant cross-section A_y (C). Applying (1) to the region (A) gives an expression for the voltage drop in this region

$$V_A = \int_0^{W_j + W_d} E_y dy = \frac{I_D(W_j + W_d)}{W(L_{diff} q N_d \mu_n) - I_D/E_C} \quad (2)$$

where the field E_y has been solved from (1), W is the width of the transistor, and N_d is the donor concentration in the drift region. The rest of the parameters have been defined in Fig. 1. On the other hand, the width of the depletion region can be expressed as a function of voltage V_A

$$W_d = \left[\frac{2\epsilon_{Si}}{qN_d} (V_{bi} + V_A + V_{ch}) \right]^{1/2} \quad (3)$$

where V_{ch} is the channel drop and V_{bi} the built-in voltage of the body pn-junction. The voltage drop V_A can be solved from (2) and (3), and it is given by

$$V_A = aW_j + \frac{a^2 b}{2} + \sqrt{a^3 W_j b + \frac{a^4 b^2}{4} + a^2 b(V_{bi} + V_{ch})} \quad (4)$$

with

$$a = \frac{I_D}{WL_{diff} q N_d \mu_n - I_D/E_C} \quad (5)$$

$$b = \frac{2\epsilon_{Si}}{qN_d} \quad (6)$$

In the region B the cross-section A_y depends on the coordinate y (the direction downwards from the surface of the

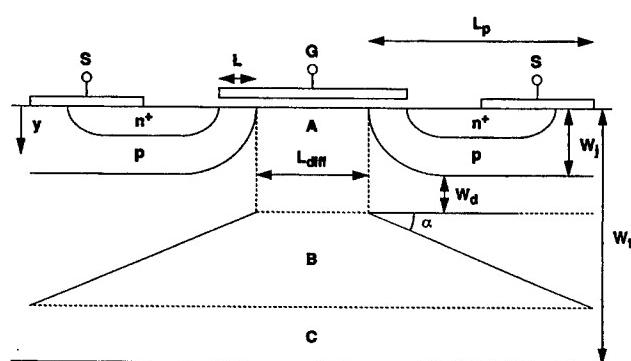


Fig. 1. Cross-section of a vertical DMOST.

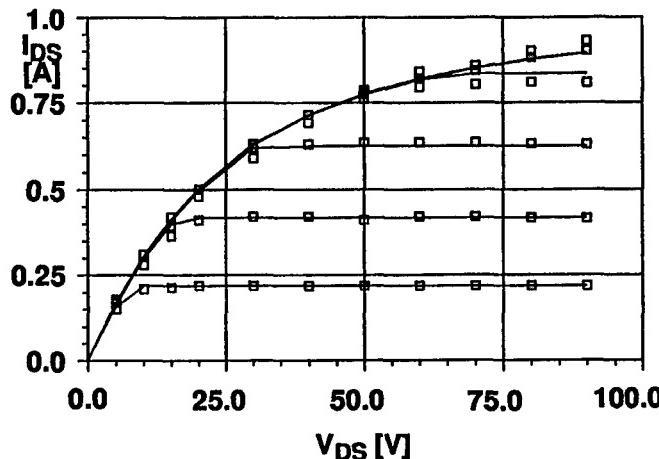


Fig. 2. Measured [2] (squares) and simulated (solid curves) I-V characteristics of a vertical power DMOST.

device)

$$A_y = W[L_{diff} + (y - W_j - W_d) \cot \alpha] \quad (7)$$

Again, by solving E_y from (1) and integrating E_y over the region (B), now taking into account (7), one obtains the voltage drop in the region (B)

$$V_B = \frac{I_D}{WqN_d\mu_n \cot \alpha} \log \left[\frac{WqN_d\mu_n(L_{diff} + L_p) - I_D/E_C}{WqN_dL_{diff}\mu_n - I_D/E_C} \right] \quad (8)$$

The same procedure results in an expression for the voltage drop in the region (C)

$$V_C = \frac{I_D(W_t - W_j - W_d - L_p \tan \alpha)}{W(L_{diff} + L_p)qN_d\mu_n - I_D/E_C} \quad (9)$$

In the channel region we can apply a short channel MOSFET model which we have published previously [3]. The model, which is based on the SPICE short channel MOS3 model [1], takes into account various short channel effects such as the electron mobility degradation due to large electric fields in the channel. In the case of the DMOST the acceptor doping concentration $N_A(x)$ has a strong gradient in the direction between source and the drift region. In accurate physical modelling this gradient has to be taken into account since it changes the threshold voltage V_T . Assuming an exponential position dependence, $N_A(x) = N_A^0 \exp(-\eta x/L)$, and following the ordinary procedure [1] for deriving the basic MOSFET equations, we get

$$V_T = V_{FB} + PHI + \frac{2}{\eta} (1 - e^{-\eta/2}) \frac{\sqrt{2q\epsilon_{Si}N_A^0}}{C_{ox}} \times F_S \sqrt{PHI - V_{BS}} \quad (10)$$

where V_{FB} , PHI and F_S are the standard SPICE parameters [1], C_{ox} the oxide capacitance, and V_{BS} the substrate bias voltage. Equation (10) reduces to the one derived previously by Kim and Fossum [2] for small values of the parameter $\eta \ll 1$. However, in real DMOST structures η is typically much larger than unity. Thus (10) is a more realistic extension of the result given in [2].

3. Model verification

The present model for vertical DMOS power transistors has been implemented in the circuit simulator APLAC [4]. The simulation results have been compared to the measured ones in Fig. 2. The experimental data was taken from [2] for a vertical DMOST structure. Figure 2 shows that the model describes very accurately the measured data. The average error is only 2.1%, which is smaller than in the K-F model [2]. So, although we neglected the space charge limited current (SCLC) in the modelling of the accumulation region (A), our model gives an excellent agreement. The dropping out of the SCLC decreases strongly the computing time needed for the simulations, since now the time consuming iterations required in the calculation of the SCLC, can be avoided.

4. Conclusions

An analytical but still device physics based model for vertical power DMOSTs has been developed. The model is more simple than the K-F model [2], as it neglects the SCLC in the modelling of the accumulation region. However, we have shown, that if the rest of the DMOST structure is treated strictly according to device physics, this more than compensates the neglect of the SCLC, resulting in a very small error between the measured and simulated results. Also the computing time needed in the simulations is thereby reduced significantly.

References

1. Antognetti, P. and Massobrio, G., "Semiconductor Device Modelling with SPICE" (McGraw-Hill, New York 1988).
2. Kim, Y-S. and Fossum, J. G., IEEE Trans. Electron Devices 37, 797 (1990).
3. Kuivalainen, P., Andersson, M., Eränen, S. and Ronkainen, H., Physica Scripta 46, 476 (1992).
4. Valtonen, M. et al., "APLAC, An Object-Oriented Analog Circuit Simulator and Design Tool", 6.2 User's Manual & Reference Manual, (Helsinki University of Technology, Circuit Theory Laboratory & Nokia Research Center, Hardware Design Technology), March 1994.

A Semi-Analytic Model of the Permeable Base Transistor

Hans-Erik Nilsson,^{1,2} Ulf Sannemo² and C. Sture Petersson¹¹ Department of Solid State Electronics, The Royal Institute of Technology, Electrum 229, S-164 28 Kista, Sweden² School of Science and Engineering, Mid-Sweden University, S-85170 Sundsvall, Sweden

Received June 6, 1994; accepted June 14, 1994

Abstract

The I - V characteristics of the Permeable Base Transistor (PBT) has been investigated in order to find a simple and practical model for use in circuit designs. Two possible approaches has been discussed, a one-dimensional analytical solution and a semi-analytical solution mixing analytical and empirical methods. The semi-analytical model developed in this paper offers high accuracy and a simple and fast evaluation. All model parameters can be extracted from a set of I - V curves from two typical transistors with different threshold voltages. An analytical small signal model has been developed that agrees very well with two-dimensional simulations.

List of symbols

μ_0	Low field mobility
$Q(V, y)$	General symbol for the channel charge
v_{sat}	Velocity saturation for carriers
V	Potential distribution in the device
N_d	Doping concentration
Z	Device dimension in z-direction
q	Carrier charge
$2a$	Grid spacing
$h(V, y)$	The depletion boundary in the channel
I_{ds}	Drain to source current
V_{ds}	Drain to source voltage
V_{gs}	Gate to source voltage
V_{gd}	Gate to drain voltage
V_T	Threshold voltage
V_c	Characteristic voltage
V_0	Contact potential
β	Empirical parameter for the mobility model
g_{ch}	Channel conductance in saturation $\frac{\partial I_{\text{ds}}}{\partial V_{\text{ds}}} \Big _{\text{saturation}}$
$I_{\text{sat}0}$	Zero biased equivalent saturation current
λ	Channel concentration modulation parameter
$I_{\text{sat}*}$	True minimal saturation current
V_{sat}	Saturation voltage
κ	Linear factor between V_{sat} and (V_0)
W	General symbol for depletion width
a_e	Equivalent radius
h_a	Depletion width towards the channel at $y = 0$
η	Empirical constant for modelling of the two-dimensional depletion
g_{linear}	Channel conductance in the linear region $\frac{\partial I_{\text{ds}}}{\partial V_{\text{ds}}} \Big _{\text{linear}}$
a_{ref}	Half the grid spacing for a reference transistor
$P(V)$	Reference polynomial
$g_{\text{ch, ref}}$	Function describing g_{ch} for a reference transistor
c_I	Empirical parameter describing the relation between g_{ch} and the grid spacing

h_D	Depletion width towards the drain
h_S	Depletion width towards the source
α_S, α_D	Shape factors
γ	The ratio between V_{ds} and $V_0 - V_{\text{gs}}$
Q_{tot}	Total depletion charge
C_{tot}	Depletion capacitance at the gate
C_{ch}	Depletion capacitance towards the channel
C_D	Depletion capacitance towards the drain
C_S	Depletion capacitance towards the source
C_{gs}	Gate to source capacitance
C_{gd}	Gate to drain capacitance
f_T	Unity current gain frequency
g_m	Transconductance
$g_{m, \text{sat}}$	Transconductance in saturation

1. Introduction

The Permeable Base Transistor was first suggested by Bozler *et al.* in 1979 [1], but it is only during the last decade the fabrication technology allows high performance PBTs. A normally on PBT biased in its ordinary region of operation is a unipolar device. The basic operation is similar to that of a short channel vertical MESFET, but the vertical orientation makes dimensions and proportions special. As the gate length often is chosen below $0.1 \mu\text{m}$, the effective channel length is mainly controlled by the depletion width towards drain and source. Several authors have examined how the unity current gain frequency f_T depends on geometry, doping profiles and semiconductor materials [2, 3]. Although simulations indicate an excellent high frequency potential, real devices are still limited by the fabrication process. Large effort has been done in order to enhance the fabrication and to reach even smaller dimensions than before. As the PBT has its strongest potential in high frequency applications small signal models have been under investigation in several papers [4, 5]. The large f_T value for a PBT makes it suitable to serve as the main active amplifying component in amplifiers with large bandwidths. However, there are other important features that also have to be considered if the component should be used as an active building block in active loads, current mirrors, digital logic and other similar applications. The DC characteristic of the PBT is not fully understood and has to be solved if the component should be used at its optimal potential. For system designs there has to be simple and accurate models available so that the designs can be simulated before fabrication. A full two-dimensional drift-diffusion simulation is too complex for the computer power available today and not practical for use in verification of larger designs.

The paper is organized as follows. In Section 2 a one-dimensional analytical model is presented, Section 3 describes the semi-analytical model, Section 4 contains an analytical small signal model of the PBT. Conclusions are given in Section 5. The software package used for two-dimensional simulation is MEDICI ver. 1.1.

2. One-dimensional model

The most commonly used MESFET models assumes: majority carrier transport, constant carrier concentration in the channel and a one-dimensional current flow. The same approximations applied to the PBT, using the mobility model proposed in [6], gives:

$$I_{ds} = \frac{\mu_0 Z Q(V, y) \frac{dV}{dy}}{\left[1 + \left(\frac{\mu_0}{v_{sat}} \frac{dV}{dy} \right)^{\beta} \right]^{1/\beta}} \quad (1a)$$

$$Q(V, y) = qN_d[a - h(V, y)] \quad (1b)$$

One critical part in equation (b) is $h(V, y)$ (see Fig. 1) since the dimension of the grid is in the same order as the depletion width. This situation makes the simple abrupt diode expression invalid since the nature of the problem is two-dimensional. Each grid line in the PBT can be looked upon as an infinite long conducting strip. Using a conformal representation the strip can be converted to an equivalent conducting cylinder. The solution of Poisson's equation for a conducting cylinder in a semiconductor environment is:

$$V_0 - V_{gs} = \frac{qN_d}{2\epsilon_0 \epsilon_r} \left[\frac{W^2}{2} - \frac{a_e^2}{2} - W^2 \log \left(\frac{W}{a_e} \right) \right] \quad (2)$$

where W is the depletion width and a_e is the cylinder radius.

For large negative V_{gs} the depletion region around a conducting strip will be a circle, corresponding to the depletion region for a certain cylinder with radius a_e . For small V_{gs} the depletion region will follow the ordinary one-

dimensional expression. A good empirical approximation of the depletion width towards the channel can be found from equation (2) as:

$$V_0 - V_{gs} = \frac{1}{\eta} \frac{qN_d}{2\epsilon_0 \epsilon_r} \left[\frac{h_a^2}{2} - \frac{a_e^2}{2} - h_a^2 \log \left(\frac{h_a}{a_e} \right) \right] \quad (3)$$

where η is an empirical constant only dependent on the grid thickness.

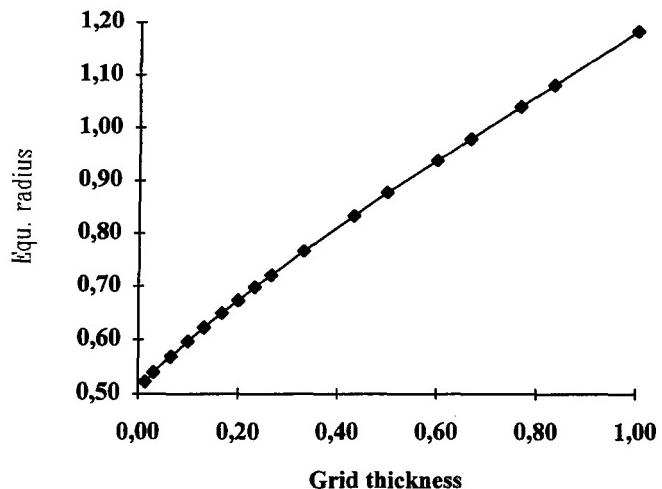


Fig. 2. Equivalent radius for normalized grid.

A diagram over equivalent radius for different grid dimensions can be seen in Fig. 2. A less accurate model proposed in [7] is to use a scaled version of the one-dimensional expression for the depletion width:

$$h_a = \sqrt{\eta \frac{2\epsilon_0 \epsilon_r (V_0 - V_{gs})}{qN_d}} \quad (4)$$

where η is an empirical constant, but not identical with the empirical constant in expression (3). The numerical value of η is 0.5 for a grid thickness of 0.02 μm and $a = 0.4$ μm.

The empirical parameter η has a very weak doping dependency and can be regarded as only dependent on grid thickness. For low drain to source voltages the depletion is almost independent of the potential distribution in the channel and depends only on the terminal voltages. Two-dimensional simulation reveals that the shape of the depletion region can be approximated as:

$$\begin{aligned} h(y) &= a + h_a \sqrt{1 - \left(\frac{y}{h_s} \right)^2}, & 0 < y \leq h_s \\ h(y) &= a + h_a \sqrt{1 - \left(\frac{y}{h_s} \right)^2}, & -h_D \leq y \leq 0 \\ h(y) &= 0, & y > h_s, \quad y < -h_D \end{aligned} \quad (5)$$

Using this approximation and solving eq. (1) with $\beta = 1$ gives:

$$\begin{aligned} \frac{V_{ds}}{\sqrt{V_0 - V_{gs}} (\sqrt{1 + \gamma} - 1)} \\ = I_{ds} A_1 \left[A_4 + \frac{\sqrt{A_2 - 1} [2 \arctan (A_2 / \sqrt{1 - A_2^2}) - \pi \sqrt{1 - A_2^2}] - i\pi \sqrt{1 - A_2}}{2A_3 A_2 \sqrt{A_2 - 1} \sqrt{1 - A_2^2}} \right] \end{aligned} \quad (6)$$

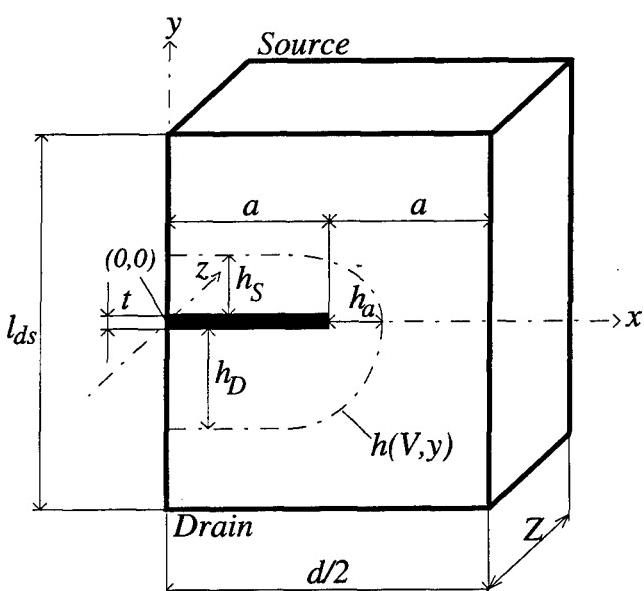


Fig. 1. Model of the PBT.

where

$$\gamma = \frac{V_{ds}}{V_0 - V_{gs}},$$

$$A_1 = \sqrt{\frac{2\epsilon_0 \epsilon_r}{qN_d}},$$

$$A_2 = \frac{v_{sat} Z q N_d A_1 \sqrt{\eta(V_0 - V_{gs})}}{v_{sat} Z q N_d a - I_{ds}},$$

$$A_3 = \mu_0 Z q N_d a - \frac{I_{ds} \mu_0}{v_{sat}},$$

$$A_4 = \frac{i\pi}{A_3 A_2 \sqrt{A_2^2 - 1}}$$

A comparison between (6) and results from a two-dimensional simulation is plotted in Fig. 3. The large error in the linear region is a direct consequence of the one-dimensional approximation.

3. Semi-analytical model

An empirical model can, in many cases, be more accurate than an analytical model, since the complexity of the analytical approach often calls for approximations that will reduce the initial accuracy considerably. On the other hand device physics is often hidden in empirical models. In spite of this limitation, empirical models have been used in many simulation tools mostly due to their efficiency when it comes to computer power and simulation time. Even when an analytical solution can be found, the non-ideal reality will often put a limit on its use, which forces us to use the analytical model in an empirical way.

The saturation of the drain current is directly related to the velocity saturation in the channel. The mobility model proposed in [6] gives the following current expression:

$$I_{ds} = \frac{V_{ds}}{V_c \left[1 + \left(\frac{V_{ds}}{V_c} \right)^{\beta} \right]^{1/\beta}} f(V_{ds}), \quad (7)$$

where $\beta = 2$ for n-type silicon.

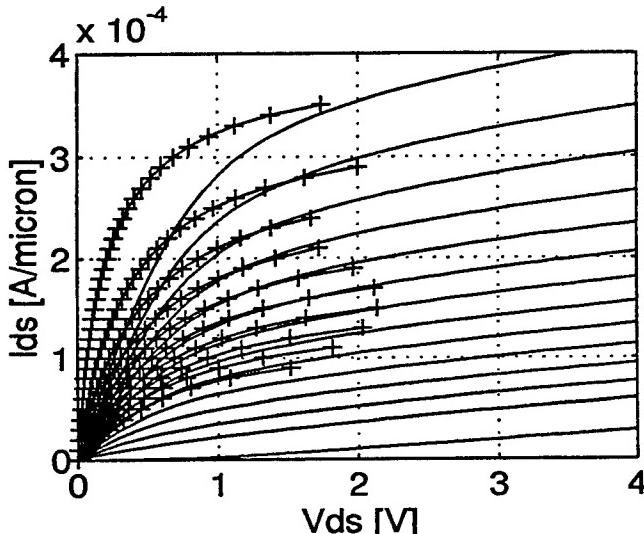


Fig. 3. Comparison between the one-dimensional model and two-dimensional simulations. (+) one-dimensional model, (Solid line) two-dimensional simulation. V_{gs} starts at 0.0 V and decreases with -1.0 V for each IV curve. Transistor dimensions are $a = 0.3 \mu\text{m}$, $t = 60 \text{ nm}$, $I_{ds} = 0.9 \mu\text{m}$.

One of the nicest characteristics of the first part in (7) is the asymptote for large drain voltages, which makes the choice of $f(V_{ds})$ very simple. The final ansatz becomes:

$$I_{ds} = \frac{V_{ds}}{V_c \left[1 + \left(\frac{V_{ds}}{V_c} \right)^{\beta} \right]^{1/\beta}} (g_{ch} V_{ds} + I_{sat0}) \quad (8)$$

V_c is a characteristic voltage which determines the transition between linear and saturation regions. The parameter g_{ch} is the channel conductance in saturation. I_{sat0} is the zero biased equivalent saturation current for $V_{ds} = 0$. In reality there is no saturation for $V_{ds} = 0$, but this is just a fictive name on the intersection between the current axis and the saturation current asymptote. The parameter β is a direct consequence of the original curve fitting parameter used for the mobility model [6]. An iterative least square method was used to fit equation (8) to results from two-dimensional simulations. The extracted parameters can be seen in Fig. 4.

The simplest parameter to view in a physical sense is I_{sat0} , as it is closely related to the minimal saturation current I_{sat*} .

$$I_{sat*} = q N_D Z (a - h_a) v_{sat}, \quad (9)$$

where h_a is the maximum depletion width in the channel, v_{sat} is the saturation velocity for electrons.

The difference is not large and can be neglected if the channel conductance is small. For large g_{ch} we have to take into account the saturation voltage V_{sat} which is related to I_{sat0} as:

$$I_{sat0} = I_{sat*} - g_{ch} V_{sat}, \quad V_{sat} \approx \kappa V_c, \quad (10)$$

where κ is independent of the grid spacing and can be extracted from a reference transistor.

The parameter V_c controls the transition between linear and saturation region. If $\beta = 2$ then V_c is the drain to source voltage where the velocity in the channel is equal to $v_{sat}/\sqrt{2}$. In the linear region the current increases due to increasing velocity. At $V_{ds} = V_{sat}$ the increase in current is no longer

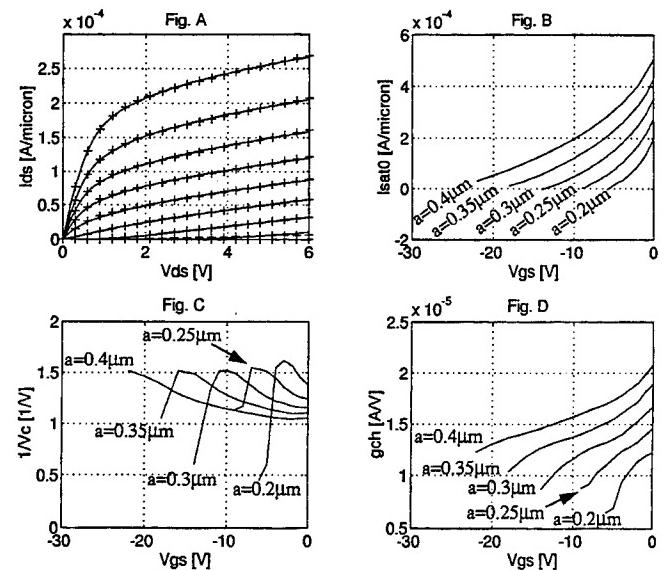


Fig. 4. (a) Semi-analytical model fitted to data from two-dimensional simulation (+) semi-analytical model, (solid line) two-dimensional simulation. V_{gs} starts at 0.0 V and decreases with -1.0 V for each IV curve. Transistor dimensions are $a = 0.2 \mu\text{m}$, $t = 60 \text{ nm}$, $I_{ds} = 0.9 \mu\text{m}$. (b) I_{sat0} as a function of gate biasing. (c) The inverse of V_c as a function of gate biasing. (d) Channel conductance g_{ch} as a function of gate biasing.

dominated by increasing velocity but rather due to increasing carrier concentration in the channel. From (8) we can see that the channel conductance for small V_{ds} (in the linear reagion) is:

$$g_{\text{linear}} = \frac{I_{\text{sat0}}}{V_c} \Rightarrow I_{ds, \text{linear}} = \frac{I_{\text{sat0}}}{V_c} V_{ds} \quad (11)$$

The linear channel conductance g_{linear} only depends on gate biasing, doping and grid dimensions. Let us assume that the grid can be modelled as a very thin strip and that the depletion regions does not reach the drain and source contacts. If these assumptions are valid we can scale the whole device with a scaling factor, including the depletion regions, and we will find that g_{linear} has been scaled with the same scaling factor. This means that if the transistor dimensions are scaled with a factor α , then the gate bias has to be scaled with a factor α^2 . Using this scaling technique we can derive g_{linear} for any PBT as long as we have g_{linear} for one transistor. As the grid finger width does not influence the current in the device a change in the grid spacing will have the same effect as a device scaling. In order to make the scaling procedure in an efficient way we need an empirical model for g_{linear} describing the gate voltage dependency. A simple and fast solution is to use a polynomial $P(V_{gs})$. V_c can be found from (11) as:

$$V_c = \frac{I_{\text{sat0}}}{\frac{a}{a_{\text{ref}}} P\left(V_{gs} \frac{a_{\text{ref}}^2}{a^2}\right)} \quad (12)$$

where a_{ref} is the grid spacing for the reference transistor and $P(V)$ is a polynomial describing the voltage dependency of g_{linear} for the reference transistor.

The same scaling procedure can be used to take into account the doping dependency. The only difference is that the scaling should be linear and not a square law [7].

The mechanism behind g_{ch} is the same as in a short channel MESFET and has been discussed at length in [8]. An analytical solution for g_{ch} is difficult to find, especially if the depletion region is disturbed by the physical boundary of the device. A simple and practical method is a linear combination of the variables most essential for a circuit for a circuit design:

$$g_{\text{ch}} = g_{\text{ch, ref}}(V_{gs}^*) + c_1(a - a_{\text{ref}}), \quad (13)$$

where c_1 is an empirical parameter independent of transistor biasing and dimensions. $g_{\text{ch, ref}}$ is a suitable empirical function describing g_{ch} for a reference transistor with grid spacing a_{ref} and V_{gs}^* is defined as:

$$V_{gs}^* = \left(\frac{a_{\text{ref}}}{a}\right)^2 V_{gs} \quad (14)$$

We need two transistors with different grid spacing in order to extract parameters for g_{ch} . A simple first order polynomial is often sufficient for $g_{\text{ch, ref}}$. An even simpler solution is to use a constant channel modulation parameter λ as in the simplest MOS-transistor model in SPICE.

$$g_{\text{ch}} = \lambda I_{\text{sat0}} \quad (15)$$

Figure 5 shows a comparison between two-dimensional simulation and the semi-analytical model for several normally on PBTs. We have used a transistor with $a = 0.3 \mu\text{m}$ to extract reference functions for g_{linear} and g_{ch} as it is

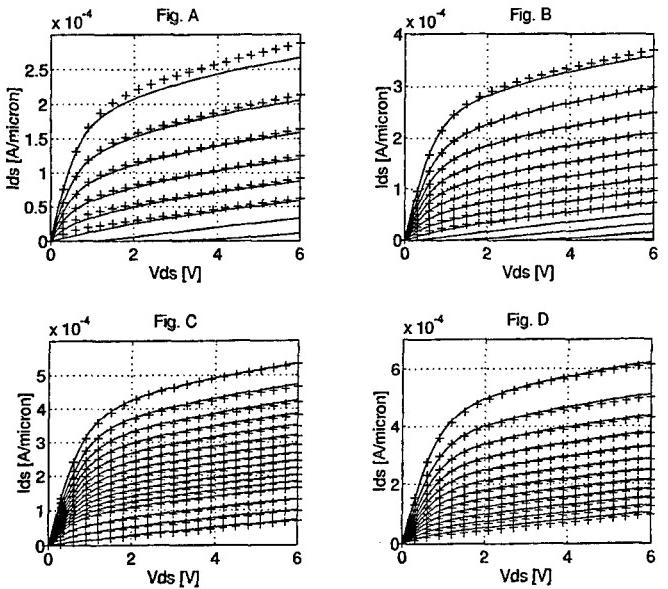


Fig. 5. (+) semi-analytical model, (solid line) two-dimensional simulation. V_{gs} starts at 0.0 V in figure (a), (b) and (c) and decreases with -1.0 V for each IV curve. In (d) V_{gs} starts at 0.0 V and decreases with -2.0 V for each IV curve. The transistor dimensions are (a) $a = 0.2 \mu\text{m}$, $t = 60 \text{ nm}$, $l_{ds} = 0.9 \mu\text{m}$. (b) $a = 0.25 \mu\text{m}$, $t = 60 \text{ nm}$, $l_{ds} = 0.9 \mu\text{m}$. (c) $a = 0.35 \mu\text{m}$, $t = 60 \text{ nm}$, $l_{ds} = 0.9 \mu\text{m}$. (d) $a = 0.4 \mu\text{m}$, $t = 60 \text{ nm}$, $l_{ds} = 0.9 \mu\text{m}$.

described in (12) and (13). The constant c_1 in (13) was extracted using g_{ch} at $V_{gs} = 0 \text{ V}$ for the transistor with $a = 0.35 \mu\text{m}$. κ in equation (10) was extracted from the reference transistor and was found to be 3.0. Note that the difference in threshold voltage between the transistor with $a = 0.4 \mu\text{m}$ and the transistor with $a = 0.2 \mu\text{m}$ is almost 15 V. For transistors with similar threshold voltage the approximation used in equation (12) and (13) should work even better.

4. Small signal model of the PBT

A normally on PBT can be regarded as a vertical MESFET. A simple small signal equivalent circuit of the PBT can be seen in Fig. 6. We have focused on the most important small signal parameters and neglected second order effects. The largest capacitance in the transistor is the depletion capacitance at the gate. The total charge in the depletion region can be modelled as:

$$Q_{\text{tot}} = Q_D + Q_S + Q_{\text{ch}} = qN_d Z A_{\text{depletion}} \quad (16)$$

$$\begin{aligned} A_{\text{depletion}} &\approx a(h_S + h_D) + h_S \int_0^{h_a} \left[1 - \left(\frac{x}{h_a} \right)^n \right]^{1/n} dx \\ &+ h_D \int_0^{h_a} \left[1 - \left(\frac{x}{h_a} \right)^n \right]^{1/n} dx \\ &= a(h_S + h_D) + h_S h_a \int_0^1 [1 - (x')^n]^{1/n} dx' \\ &+ h_D h_a \int_0^1 [1 - (x'')^n]^{1/n} dx'' \\ &= a(h_S + h_D) + h_S h_a \alpha_S + h_D h_a \alpha_D \end{aligned} \quad (17)$$

$$\begin{aligned} Q_{\text{tot}} &= Q_D + Q_S + Q_{\text{ch}} \\ &= qN_d Z [a(h_D + h_S) + \sqrt{\eta} \alpha_S h_a^2 + \sqrt{\eta} \alpha_D h_D h_a] \end{aligned} \quad (18)$$

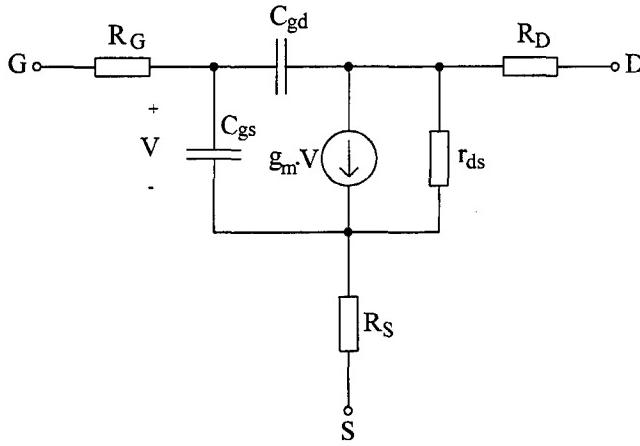


Fig. 6. Small signal model of the PBT.

According to (17) the value of α_D and α_S should be somewhere between 0.5 and 1.0 ($\pi/4$ for $n = 2$). Numerical values extracted from two-dimensional simulation gives $\alpha_D = 0.5$ and $\alpha_S = 0.5$. Using (18) we can write the total capacitance as:

$$\begin{aligned} C_{\text{tot}} &= \frac{\partial Q_{\text{tot}}}{\partial V_{\text{gs}}} \\ &= Z\varepsilon_0 \varepsilon_r \left(\frac{a}{h_D} + \frac{a}{h_S} + 2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \chi \right) \\ &= C_D + C_S + C_{\text{ch}}, \end{aligned}$$

where

$$C_{\text{ch}} = Z\varepsilon_0 \varepsilon_r (2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \chi)$$

$$\chi = \sqrt{1 + \gamma} \left(\frac{1}{1 + \gamma} + 1 \right), \quad \gamma = \frac{V_{\text{ds}}}{V_0 - V_{\text{gs}}}$$

$$\begin{aligned} C_{\text{gs}} &= \frac{\partial Q}{\partial V_{\text{gs}}} \Big|_{V_{\text{gd}} = \text{constant}} \\ &= Z\varepsilon_0 \varepsilon_r \left(\frac{a}{h_S} + 2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \sqrt{1 + \gamma} \right) \end{aligned}$$

$$\begin{aligned} C_{\text{gd}} &= \frac{\partial Q}{\partial V_{\text{gd}}} \Big|_{V_{\text{gs}} = \text{constant}} \\ &= Z\varepsilon_0 \varepsilon_r \left(\frac{a}{h_D} + 2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \frac{1}{\sqrt{1 + \gamma}} \right) \end{aligned}$$

$$C_{\text{tot}} = C_{\text{gd}} + C_{\text{gs}}$$

If C_D and C_S is large in comparison with C_{ch} we can regard C_{ch} as a constant. This approximation will be very accurate for large $V_0 - V_{\text{gs}}$ (Fig. 7).

The transconductance in saturation can be expressed as:

$$\begin{aligned} g_{m, \text{sat}} &= \frac{\partial I_{\text{sat0}}}{\partial V_{\text{gs}}} + V_{\text{ds}} \frac{\partial g_{\text{ch}}}{\partial V_{\text{gs}}} \\ &\approx \frac{Zv_{\text{sat}} \eta \varepsilon}{h_a} + V_{\text{ds}} \frac{\partial g_{\text{ch, ref}}}{\partial V_{\text{gs}}^*} \left(\frac{a_{\text{ref}}}{a} \right)^2 \end{aligned} \quad (22)$$

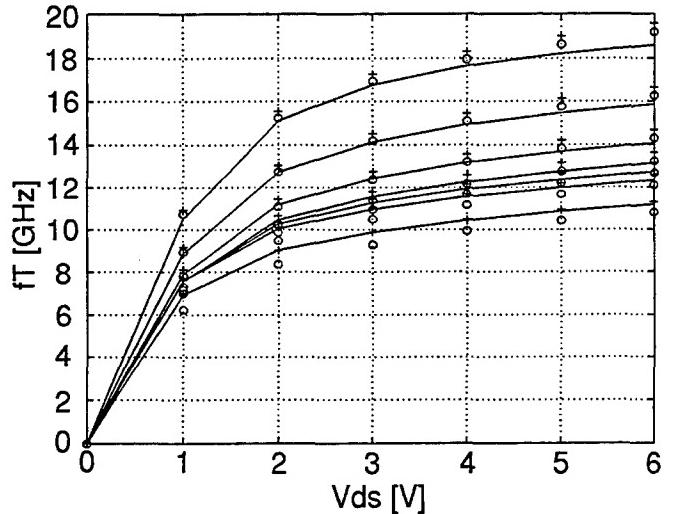


Fig. 7. Comparison between f_T extracted from two-dimensional simulation and the analytical model. (Solid line) two-dimensional simulation, (○) analytical model, (+) reduced model ($C_{\text{ch}} = \text{constant}$). Transistor dimensions are $a = 0.3 \mu\text{m}$, $t = 60 \text{ nm}$, $l_{\text{ds}} = 0.9 \mu\text{m}$.

Using (18) and (21) we can write the unity current gain frequency f_T for the saturation region as:

$$\begin{aligned} f_T &= \frac{g_{m, \text{sat}}}{2\pi C_{\text{tot}}} \approx \frac{v_{\text{sat}} \eta}{2\pi h_a \left(\frac{a}{h_D} + \frac{a}{h_S} + 2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \chi \right)} = \frac{v_{\text{sat}}}{2\pi L_{\text{eff}}} \\ L_{\text{eff}} &= \frac{h_a}{\eta} \left(\frac{a}{h_D} + \frac{a}{h_S} + 2\sqrt{\eta} \alpha_S + \sqrt{\eta} \alpha_D \chi \right) \end{aligned} \quad (23)$$

It can be seen in (19), (22) and (23) that for high speed the doping level should be as high as possible and the gate biasing should be chosen to minimize the depletion regions. The transconductance will increase and the total capacitance will decrease for decreasing grid thickness resulting in a higher f_T . In Fig. 7 we have used expression (19) for the capacitance and the full I-V model has been numerically differentiated in order to have g_m for both linear and saturation region.

It is interesting to see that a limiting value on f_T for the drift-diffusion model can be estimated as:

$$\lim_{a \rightarrow 0} f_T \propto \frac{v_{\text{sat}} \sqrt{\eta}}{2\pi h_a (2\alpha_S + \alpha_D \chi)} \approx 80 \text{ GHz}, \quad \alpha_D = \alpha_S = 0.5, \quad \chi = 3, \quad \eta = 0.5 \quad (24)$$

for $N_d = 10^{17} \text{ cm}^{-3}$. In (24) the variable a stands for the finger width and not for grid spacing. In this limiting case the PBT is a vertical MESFET with extremely short channel. For these small dimensions hot electron effects are expected to be important, which makes (24) a conservative estimation.

5. Conclusions

A one-dimensional model of the PBT in MESFET operation has been presented and compared with two-dimensional simulation. The comparison shows that two-dimensional effects has to be considered in an accurate I-V model. A semi-analytical model including two-dimensional effects has been presented. The model consists of a set of parameters that can be extracted from two refer-

ence transistors. The model shows good agreement with two-dimensional simulations. We have chosen to compare the model with two-dimensional simulations, even though it has no problem to fit measured data as well.

An analytical expression for f_T has been proposed which agrees very well with results from two-dimensional simulations and explains the simulation results published by Bozler and Alley [2] concerning the f_T dependency on device geometry and voltage biasing. The small signal analysis is based on the drift-diffusion model and does not include any hot electron effects. For small devices other transport models has to be used. However, it is not clear how these effects will influence the device and further research in low level modelling is needed before a high level model can be found.

References

1. Bozler, C. O., Alley, G. D., Murphy, R. A., Flanders, D. C. and Lindley, W. T., Proc. 7th Bien. Cornell Conf. on Active Microwave Semicond. Dev., p. 33 (1979).
2. Bozler, C. O. and Alley, G. D., IEEE Trans. Electron Devices, Vol. ED-21(6), (1980).
3. Rathmann, D., IEEE Trans. Electron Devices, 37(9), (1990).
4. Vojac, B. A. and Alley, G. D., IEEE Trans. Electron Devices, ED-24(8), (1983).
5. Matthews, D. S., Pitzalis, O., Jr. and Lee, R. E., Proc. IEEE/Cornell Conf. on High-Speed Semiconductor Devices and Circuits (1983).
6. Thornber, K.K., J. Appl. Phys., 51, 2121 (1980).
7. Cheneview, P., Kamarinos, G. and Pananakais, G., Proc. ESDERC 93, 121 (1993).
8. Sze, S. M., "Physics of Semiconductor Devices" (John Wiley & Sons, Inc. 1981), ch. 6, p. 334.

Interferometric, Low Thermal Mass IR-absorber for Thermal Infrared Detectors

P. Eriksson, J. Y. Andersson* and G. Stemme

Instrumentation Laboratory, Royal Institute of Technology (KTH), Stockholm, Sweden

Received May 9, 1994; accepted June 15, 1994

Abstract

The absorption efficiency of thermal infrared detectors of various designs is theoretically evaluated and compared, with an emphasis on detectors possible to fabricate by surface micromachining. In particular, the requirement of low thermal mass is considered. An absorber consisting of a single resistive metal film is shown to give a maximum of 50% absorptance. By backing such a film with a perfect reflector located at a $\lambda/4$ distance from it, nearly 100% absorptance can be attained in the broad wavelength range of 8–12 μm . The simulations show that by adding a dielectric layer (membrane) onto the metal film, the absorptance remains nearly constant, provided that the proper values of n_d , d_1 , R_s and d_3 are chosen. For $d_1 = 0.2 \mu\text{m}$, the maximum variation in d_3 and R_s , compatible with a decrease in mean absorptance from 100 to 95%, is 30 and 40%, respectively. If the dielectric film possesses absorptive behaviour the absorptance of the detector may still become large for previously selected structure parameters.

1. Introduction

Infrared technology has been rapidly expanding over the past two years with great emphasis on detector arrays, including large, high resolution ($>128 \times 128$ picture elements) focal plane arrays (FPA). Recently, arrays based on thermal detectors have evolved as viable candidates for thermal imaging applications [1]. The operation of thermal detectors depends on a two-step process. The absorption of light raises the temperature of the device, which in turn changes some temperature dependent parameter such as electrical conductivity. Thermal detectors may be thermopile (Seebeck effect), resistive or dielectric bolometer, pyroelectric or Golay cell (thermopneumatic). The types most suitable for detector arrays are the bolometer and the pyroelectric types. Characteristic features of thermal detectors are the advantage of room temperature operation, but unfortunately combined with low sensitivity (signal to noise ratio) and slow response. For FPAs, however, where the modulation bandwidth is low and the ability to integrate for a frame time, implies that the latter drawbacks become less serious. In order to optimize sensitivity and speed, high absorptance has to be achieved in conjunction with a low thermal mass (heat capacity). The absorptance A is defined as the ratio of the absorbed power to the incident optical power of the detector. A commonly used measure of performance for the signal to noise ratio of an IR detector is the detectivity D^* , which can be expressed as:

$$D^* = AF \frac{SR_{th}}{u_N} \sqrt{(a_d \Delta f)} \quad (1)$$

where F is the fill factor (the ratio of the optical area and total area of the picture element), S the transfer function of the temperature sensitive material or structure (V/K), R_{th} the thermal resistance (K/W), u_N the noise voltage spectral density, a_d the detector area, and Δf the noise bandwidth. Evidently, the largest possible absorptance is crucial for obtaining a high sensitivity or D^* . In addition the thermal resistance need to be high. Considering the speed of the detector, its time constant is ultimately limited by the thermal time constant which depends on the thermal mass C_{th} according to

$$\tau_{th} = R_{th} C_{th} \quad (2)$$

A high speed detector evidently requires a low thermal mass.

Micromachining is suitable for fabrication of large thermal detector arrays of the resistive bolometer type. In particular, surface micromachining [2], which involves the processing of membrane or bridge structures, suspended clear of the underlying silicon substrate by a couple of legs (usually two or four). This method permits read-out electronics to be implemented in the substrate beneath the membrane, thus making possible the processing of arrays with large pixel fill factor. In addition, since these membranes can be made thin the thermal mass may be low. A high thermal resistance can be obtained by making the supporting legs sufficiently long and thin.

Conventional absorbers suffer from either a high thermal mass, e.g. porous films [3], or a relatively low absorptance, e.g. thin resistive metal films [4].

In order to alleviate these drawbacks we here discuss absorber concepts based on interferometric structures. A structure of this type utilizes an impedance matched metal film, backed with a perfect reflector at a $\lambda/4$ optical distance [5] (λ will henceforth refer to the wavelength in the medium). Such an absorber enables high absorptance combined with a low thermal mass, and will henceforth be referred to as IS. However, to be exploited in a bolometer, the thermo-sensitive material has to be electrically isolated from the metal by a dielectric film, which modifies the optical properties of the structure (see Fig. 1). This dielectric film, or membrane, also acts as a mechanical support for the thermo-sensitive film deposited on its upper surface, and the thin metal film absorber at its lower one. Such a modified absorber (MoIS) can favourably be implemented in a bolometer.

This article will outline necessary considerations when designing interferometric absorbers, with emphasis on the MoIS type absorbers.

* Industrial Microelectronic Center, Stockholm, Sweden.

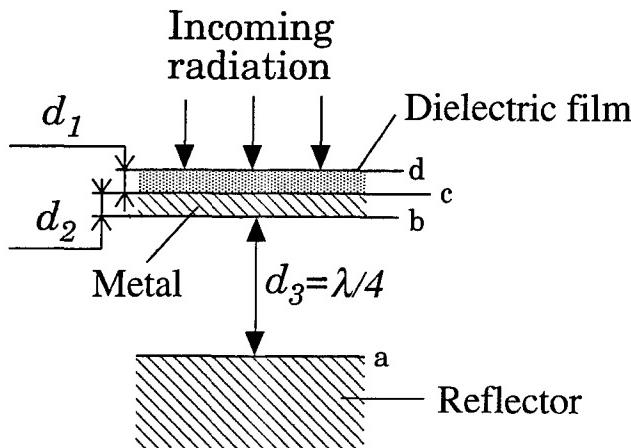


Fig. 1. A quarter wavelength structure where a dielectric film has been added on top of the metal film.

2. Theory

To look for explicit expressions for the absorptance of multiple films is cumbersome due to the tedious manipulations of complex numbers. However, by using either (i) the transfer matrix method or (ii) the impedance transformation method, such calculations are considerably simplified. In addition, both are suitable for numerical simulations.

The transfer matrix method distinguishes between reflection and transmission matrices. By using boundary conditions obtained from Maxwell's equations the reflection matrix can be derived relating the incident and the reflected waves on each side of a boundary between two media. The phase shift introduced when a plane wave is propagating through a medium is calculated from the transmission matrix. If the boundary between medium 1 and 2 is characterised by its reflection coefficient Γ_{1-2} and transmission coefficient τ_{1-2} , with the radiation incident from medium 1, the reflection matrix R_{1-2} and the transmission matrix T_1 become

$$R_{1-2} = \frac{1}{\tau_{1-2}} \begin{pmatrix} 1 & \Gamma_{1-2} \\ \Gamma_{1-2} & 1 \end{pmatrix}$$

$$T_1 = \begin{pmatrix} \exp(-jk_1 d_1) & 0 \\ 0 & \exp(jk_1 d_1) \end{pmatrix} \quad (3)$$

where k_1 and d_1 are the complex propagation constant and film thickness, respectively, of medium 1.

The impedance method is often used by microwave engineers but due to the analogy between the wave characteristics on a transmission line and plane waves in a lossy medium it is equally applicable to this kind of problem. The total wave impedance is transferred from one interface (impedance = Z_{2-3}) towards another (impedance = Z_{1-2}) in the direction of the source of radiation (see Fig. 2). The medium is assumed to have an intrinsic impedance η_2 . One obtains

$$Z_{1-2}(d) = \eta_2 \frac{Z_{2-3} - j\eta_2 \tan(k_2 d)}{\eta_2 - jZ_{2-3} \tan(k_2 d)} \quad (4)$$

The reflection coefficient at the interface 1 – 2 becomes

$$\Gamma_{1-2} = \frac{Z_{1-2} - \eta_1}{Z_{1-2} + \eta_1}. \quad (5)$$

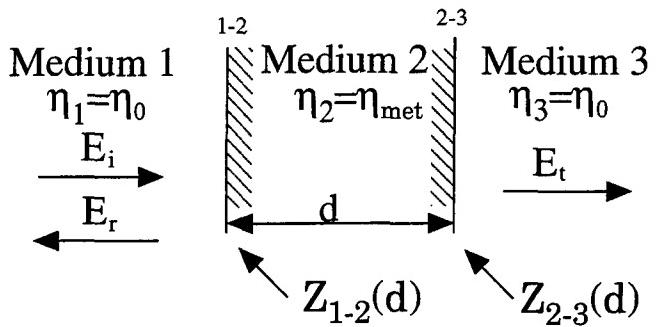


Fig. 2. A single resistive metal film of thickness d surrounded by vacuum. η_0 is the intrinsic wave impedance of vacuum.

This technique has the advantage of being more illustrative when used in conjunction with the Smith chart.

To make calculations manageable some approximations are needed. The radiation is throughout considered as plane waves propagating perpendicularly to the plane of films, i.e. transverse electromagnetic waves (TEM). The dielectric films are considered loss free, if not otherwise stated. In the far IR range (wavelength 8–12 μm) the propagation constant of a metal can be approximated as

$$k = \sqrt{\frac{\pi\sigma}{c_0 \epsilon_0 \lambda_0}} (1 + j) \quad (6)$$

where σ is the electrical conductivity, ϵ_0 the permittivity of free space and c_0 and λ_0 the speed of light and the wavelength, respectively, both in vacuum.

As an example of the impedance method, the absorptance of a single metal film can be deduced from the expression of the total wave impedance at the interface 1 – 2 (see Fig. 2) which has been transferred from interface 2 – 3. Since

$$Z_{2-3} = \eta_0 = \sqrt{\frac{\mu_0}{\epsilon_0}}, \quad (7)$$

where η_0 is the intrinsic impedance of vacuum (= 377 Ω), and μ_0 the magnetic permeability of vacuum, the transferred total wave impedance at interface 1 – 2 can be found from eq. (4) as

$$Z_{1-2}(d) = \eta_{met} \frac{\eta_0 - j\eta_{met} \tan(k_2 d)}{\eta_{met} - j\eta_0 \tan(k_2 d)} \quad (8)$$

For the thin metal film the phase change $k_2 d$ is assumed to be small and thus

$$\tan(k_2 d) \approx k_2 d. \quad (9)$$

Using eqs. (8) and (9) and simplifying one ends up with

$$Z_{1-2} \approx \frac{R_s \eta_0}{R_s + \eta_0}, \quad (10)$$

where R_s is the sheet resistivity of the metal film. A negligible change of phase angle across the film [eq. (9)] implies constant electric and magnetic fields. From this fact and the boundary conditions of the E - and H -field one obtains

$$1 + \Gamma_{1-2} \approx \tau_{1-3} \quad (11)$$

where τ_{1-3} is the transmission coefficient through the film. From

$$A = 1 - |\Gamma_{1-2}|^2 - |\tau_{1-3}|^2, \quad (12)$$

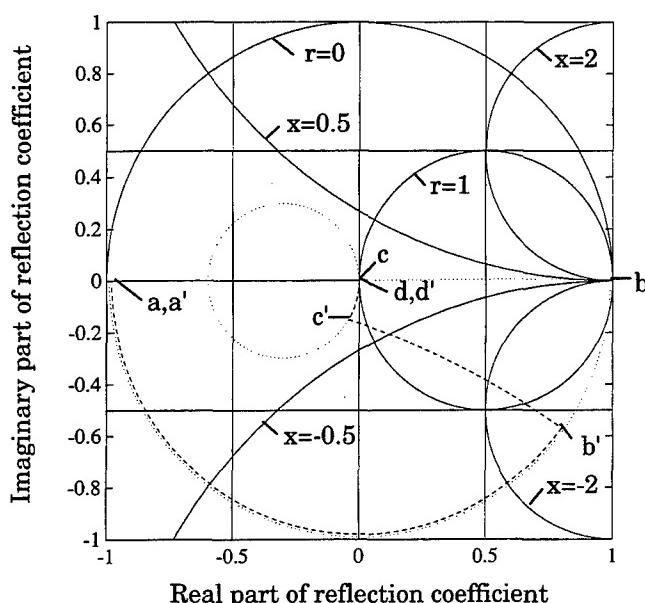


Fig. 3. The impedance transformation depicted in the Smith chart, where the normalized total wave impedance is $z = r + jx$.

the absorptance A reaches its maximum when $\Gamma_{1-2} = -0.5$, and the corresponding sheet resistivity is found to be,

$$\Gamma_{1-2} = \frac{Z_{1-2}(d) - \eta_0}{Z_{1-2}(d) + \eta_0} = -0.5 \Leftrightarrow R_s = \frac{\eta_0}{2}. \quad (13)$$

Hence, the maximum absorptance achieved by a single metal film is only 50%, obtained when $R_s = 188 \Omega$ (half the intrinsic impedance of vacuum). In a similar way it can be shown that for the IS type absorber one ends up with $A = 100\%$, provided that $R_s = \eta_0$ and 100% efficiency of the reflector. Evidently, this is a major improvement compared to the case of a single metal film.

3. Numerical simulation

The simulations have been carried out in order to find out whether the high absorptance of IS can be maintained even if a dielectric layer is present, as depicted in Fig. 1.

The simulations make use of the transfer matrix method at a wavelength of $10 \mu\text{m}$ if not otherwise stated. The refractive index of the dielectric film has been chosen as $n_d = 2$ (which is close to the refractive index of silicon nitride).

The behaviour is illustrated by the Smith chart. The impedance in the diagram has been normalized with respect to η_0 . When transferring the impedance along a-b-c-d in Fig. 1 it will follow the dotted line (a-b-c-d) in Fig. 3. When moving across the dielectric layer the reflection coefficient will follow a circular loop in a counter clockwise direction whose radius depends, among other factors, on the intrinsic impedance of the dielectric medium. One turn around the circle corresponds to an optical thickness of the layer equal to $\lambda/2$. Since the reflector is assumed to be loss free the absorption completely takes place within the thin metal film. One thus obtains for the absorptance that $A = 1 - |\Gamma|^2$ which implies that the closer origo the higher the absorptance. From this it is evident that a 100% absorptance can be attained by letting the thickness of the dielectric medium be $\lambda/2$, which for the case of $n_d = 2$ and $\lambda = 10 \mu\text{m}$ corresponds to $d_1 = 2.5 \mu\text{m}$. This extra thermal mass can be reduced by choosing a dielectric medium with a higher refractive index, but if the wavelengths of interest are $8-12 \mu\text{m}$ and the materials of the type normally used in IC production, the added thermal mass will slow down the sensor too much.

One way to circumvent the problem is to choose a smaller value than $\lambda/4$ for the air gap. When transferring the impedance in the same way along a-b-c-d with a smaller air gap, the dashed line (a'-b'-c'-d') in Fig. 3 will be followed. When entering the dielectric medium, instead of moving away from origo the reflection coefficient will move towards origo. With the correct relation between n_d , d_1 , R_s and d_3 a 100% absorptance is achieved. When two of these four parameters are known, the other two follows, i.e. the number of degrees of freedom = 2 under these assumptions. The calculated example assumes an air gap of $\lambda/5$, a sheet resistivity of 366Ω and a dielectric film thickness of 170 nm . This thickness corresponds to an acceptable value of the total thermal mass.

The freedom experienced when choosing thicknesses is valuable since other aspects have to be considered when

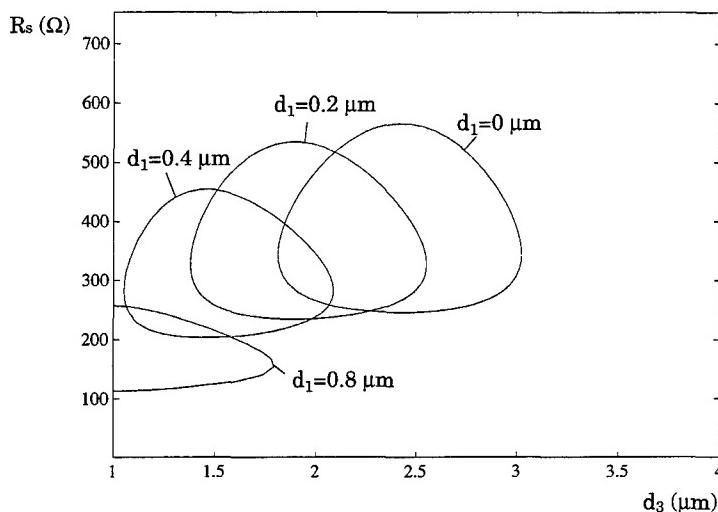


Fig. 4. Contours of 95% mean absorptance for four different values of the thickness of the dielectric film d_1 . Detector structure: see Fig. 1.

fabricating the absorber. The residual tensions left in the dielectric film after deposition may result in a deflection of the membranes, leading to the air gaps d_3 of the individual pixels of the array being difficult to reproduce. In addition, in order to control the tensions it is desirable to have a freedom in the choice of the membrane thickness.

The sensitivity of the absorptance to changes in the parameters d_1 , R_s and d_3 is shown in Fig. 4 where the contour of 95% absorptance is depicted for four different values of d_1 . In this calculation the absorptance A_m is a weighted mean value over the range 8–12 μm according to,

$$A_m = \frac{\int_8^{12} A(\lambda) W(\lambda) d\lambda}{\int_8^{12} W(\lambda) d\lambda}, \quad (14)$$

where $W(\lambda)$ is the emitted energy per unit wavelength from a blackbody. These results make evident the insensitivity to variations in the parameters even when a dielectric film is present. From Fig. 4 it is found that for $d_1 = 0.2$ μm, the maximum variation in d_3 and R_s , compatible with a decrease in mean absorptance from 100 to 95%, is 30 and 40%, respectively. This is valuable in view of the fact that a spread in the air gap d_3 is expected. The spectral absorptance is shown in Fig. 5 for the case of $d_1 = 0$ and 0.2 μm, respectively, with R_s and d_3 selected at optimum mean absorptance. It is noteworthy that the absorptance in the 8–12 μm wavelength region is nearly constant.

If the dielectric film possesses absorptive behaviour the absorptance may still become large for properly selected structures. For example, the larger the absorption in the dielectric film the larger is R_s . The latter fact is compatible with a smaller absorption in the metal, thus keeping the total absorption nearly constant. Difficulties are encountered when the absorption in the dielectric film is strongly wavelength dependent. In this case the broadband behaviour of the absorption cannot in general be maintained.

4. Summary

The absorption efficiency of thermal infrared detectors of various designs is theoretically evaluated and compared. In particular, the requirement of low thermal mass is considered. An absorber consisting of a single resistive metal

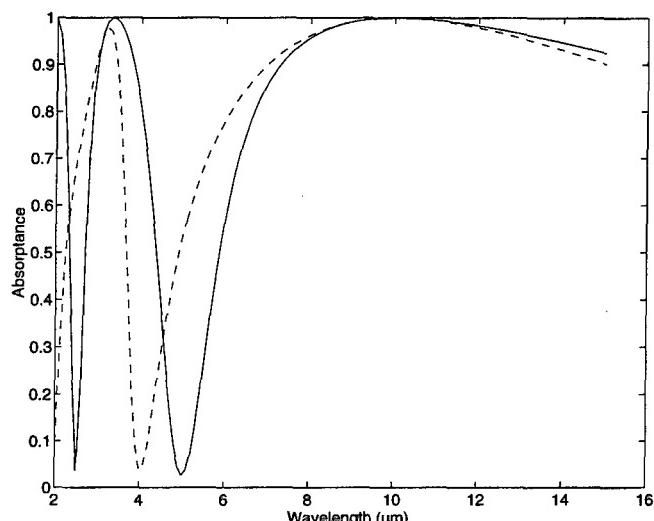


Fig. 5. The spectral absorptance vs. wavelength for a detector structure of the interferometric type. The full curve refers to the case of no dielectric layer, whereas the dashed curve assumes a thickness of the dielectric layer equal to 0.2 μm. Optimum values of sheet resistance and air gap distance are chosen.

film is shown to reach a maximum of 50% absorptance. By backing such a film with a perfect reflector located at a $\lambda/4$ distance from it, 100% absorptance can be attained in the broad wavelength range of 8–12 μm. The simulations show that by adding a dielectric layer (membrane) onto the metal film, the absorptance remains nearly constant, provided that the proper values of n_d , d_1 , R_s and d_3 are chosen. For $d_1 = 0.2$ μm, the maximum variation in d_3 and R_s , compatible with a decrease in mean absorptance from 100 to 95%, is 30 and 40%, respectively. If the dielectric film possesses absorptive behaviour the absorptance of the detector may still become large for properly selected structure parameters.

References

1. Wood, R. A., Proceeding of SPIE, **2020**, 322 (1993).
2. Howe, R. T., J. Vac. Sci. Technol. **B6(6)**, 1809 (1988).
3. Lang, L., Kühl, K. and Sandmaier, H., Sensors and Actuators **A34**, 243 (1992).
4. Woltersdorff, W., Z. Phys., **91**, 230 (1934).
5. Parsons, A. D. and Pedder, D. J., J. Vac. Sci. Technol. **A6(3)**, 1686 (1988).

UV-Sensitive Photodetectors Based on Metal-Semiconductor Contacts on 6H-SiC

Christer Fröjdः*, Göran Thungström*, Hans-Erik Nilsson* and C. Sture Petersson

Royal Institute of Technology, Dept of Solid State Electronics, Electrum 229, S-164 40 Kista, Sweden

Received June 6, 1994; accepted June 15, 1994

Abstract

Schottky diodes on Silicon Carbide (SiC) are of interest for many applications because of the relatively simple fabrication process compared to pn diodes. In this work we have fabricated Schottky diodes by evaporation of Ti on n-type and p-type 6H-SiC. Most of the diodes show good rectifying behaviour with very low reverse current and an ideality factor below 1.20. The photo response of the diodes has been measured in the range 200–400 nm. The peak sensitivity varies in the range 250–300 nm depending mainly on substrate doping.

1. Introduction

Silicon Carbide is considered as a promising material for high power and high frequency applications because of its wide bandgap ($E_g = 2.9$ eV for 6H-SiC), high saturation velocity for electrons and high thermal conductivity. A review of the properties of the material can be found in [1, 2].

The wide bandgap also makes the material interesting for measuring UV-radiation. Photodiodes made on 6H-SiC are only sensitive to radiation at wavelengths below 400 nm, the wavelength where the photon energy corresponds to the bandgap energy. Their advantage as compared to Silicon photo diodes is that they can be used to measure the UV-radiation in visible light without extra filtering.

The absorption coefficient for light in SiC is strongly wavelength dependent varying from about 10^2 to 10^6 cm⁻¹ when going from 400 nm to 200 nm. In order to make efficient photo diodes for short wavelengths the window of the diode must be very thin. The other important property is that the leakage current must be kept low.

Photodiodes with pn-junctions have already been fabricated by ion implantation [3, 4]. High quality photo diodes fabricated by mesa etching of epitaxial layers were presented in [5]. Schottky photo diodes on n-type 6H-SiC, fabricated by our group, were presented in [6].

Many metals form diodes with high Schottky barriers when deposited on silicon carbide. A number of reports on diodes made from Pt and Ti [7], Co [8] and other metals on n-type SiC exist in the literature. However very few studies of Schottky barriers on p-type SiC have been presented. One exception is [9]. The advantage in using a Schottky diode as a photodiode is the simple fabrication process and that no absorbing window exists, since the light enters through finger shaped openings in the metal. The disadvantage is of course that one part of the diode area is shielded from light by the metal.

In a previous project [6] we fabricated finger shaped Schottky diodes by evaporating Ti on a n-type 6H-SiC wafer. In this study diodes on substrates with different n-type and p-type doping levels have been investigated.

The following diodes have been fabricated:

N1. n-type 6H-SiC with no visible defects. $N_D \approx 7 * 10^{15}$ cm⁻³, thickness of epitaxial layer is 1.6 μm.

N2. n-type 6H-SiC. This sample was used in [5] and is included here for comparison. $N_D \approx 2 * 10^{16}$ cm⁻³, thickness of the epitaxial layer is about 10 μm.

N3. n-type 6H-SiC with very high doping. $N_D \approx 2 * 10^{18}$ cm⁻³, thickness of epitaxial layer is about 5 μm.

N4. n-type 6H-SiC with high density of pinholes. $N_D \approx 1 * 10^{17}$ cm⁻³, thickness of epitaxial layer is about 5 μm.

P1. p-type 6H-SiC. $N_A \approx 1 * 10^{17}$ cm⁻³, thickness of epitaxial layer is about 10 μm.

2. Processing

The starting material was 6H-SiC wafers with an epitaxial layer. The doping of the bulk wafers and the epitaxial layers are specified in Table I. The wafers were fabricated by Cree Research Inc. Prior to processing the wafers were degreased by subsequent dipping in trichloroethylene, acetone and 2-propanol and finally dipped in HF. A pattern with a number of finger shaped openings was formed on the wafer. The finger widths were in the range 1–2 μm. Four different diode areas were used: 100 × 100 μm, 75 × 75 μm, 50 × 50 μm and 25 × 25 μm. A SEM-micrograph of a part of the pattern is shown in Fig. 1. The left part of the metal area is used for probing. The total size of the metal area in

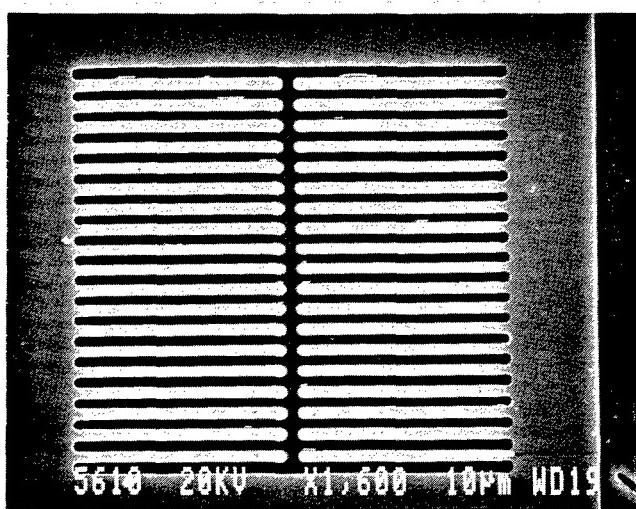


Fig. 1. SEM micrograph showing the finger area of one photodiode. To the left of the fingers is a metal area for probing. The area of the finger opening for this diode is 50 × 50 μm and the finger width is 1.2 μm.

* Also: Mid-Sweden University, Dept. of Electronics, S-851 70 Sundsvall, Sweden.

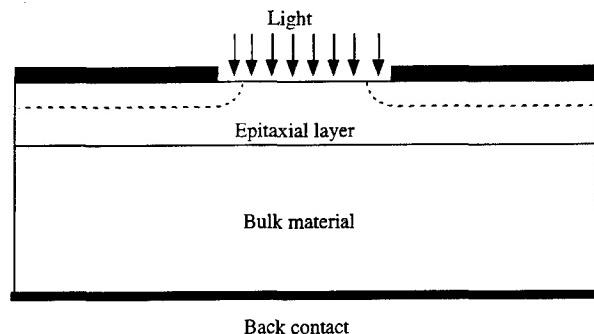


Fig. 2. Cross section of a photodiode. Thickness and doping of the epitaxial layer is specified in table I for each sample. The depletion region is indicated by the dotted line in the figure. Photons absorbed in the depletion region or within one diffusion length from the depletion region contribute to the photocurrent.

one diode is $120 \times 220 \mu\text{m}$. Each diode is surrounded by an opening with a width of $10 \mu\text{m}$. A cross section of one part of a diode is shown in Fig. 2.

Titanium was evaporated on the wafers using an electron gun evaporator with a residual pressure in the range of 10^{-8} Torr. The resulting layer thickness was 2000 \AA . A standard lift-off procedure was used to remove excess metal. A back contact was made by evaporation of 5000 \AA of Al on the back side of the wafer. After metallisation the samples were annealed in a furnace at 515°C for 30 minutes. No additional surface passivation was done.

3. Electrical characterisation

After processing the diodes were characterised electrically by IV and CV measurements. A probe tip was placed on the metal area on the left side of the opening in the diode. The back side of the wafer was used as the second electrode.

CV-measurements were performed using a HP4279A CV-meter at a frequency of 1 MHz and varying the bias voltage from 0 to -10 V . Figure 3 shows the results from the CV-measurements of all diodes. The doping concentration in the epitaxial layer and the Schottky barrier heights for the different samples are presented in Table I.

IV-characteristics were measured in a computerised measurement system using HP3245A Universal Source and HP3458A Multimeter.

Forward current as a function of forward voltage for diodes on all the different SiC substrates are plotted in Fig. 4. Most of the diodes show good rectifying behaviour with

Table I. Measured characteristics for the diodes on different SiC material. Doping is calculated from CV-measurements Barrier heights are calculated from both CV and IV measurements. Ideality factor (N) is calculated from IV-measurements

	$N_D(N_A)$	B CV	B IV	N	$I_r@ -10 \text{ V}$
N1	$7.5 * 10^{15}$	0.98	0.97	1.08	<10 pA
N2	$2.9 * 10^{16}$	0.89	0.92	1.14	0.4 nA
N3	$2.7 * 10^{18}$	0.94	0.79	1.23	0.5 mA
N4	$1.0 * 10^{17}$	0.94	0.84	1.17	18 μA
P1	$1.6 * 10^{17}$	2.43	1.94	1.03	<1 pA

an ideality factor generally below 1.20 and with low reverse leakage currents. The diodes on sample N3 also show good rectifying behaviour despite the very high doping in the semiconductor. However in this case the ideality factor and the reverse leakage current are higher than for the other diodes.

For diodes on the p-type material the reverse leakage current at -10 V was below 1 pA and for a reverse voltage of up to -70 V the leakage current remained below 500 pA. For the diodes on the n-type substrates the reverse current was below 1 nA at -10 V except for diodes on the material with the highest doping and the material with a high density of defects.

4. Optical characterisation

Optical measurements were made using an ORIEL 1/8 m single monochromator equipped with an UV-enhanced 150 W Xenon arc lamp. Relative responsivity was measured from 400 to 200 nm in steps of 10 nm with a spectral bandwidth of 10 nm. The total beam area during this measurement was $3 \times 3 \text{ mm}$ and the intensity of the light varied from 1 mW at 400 nm down to $12 \mu\text{W}$ at 200 nm. The photo response for diodes with an area of $100 \times 100 \mu\text{m}$ and a finger width of $2 \mu\text{m}$, measured at 0 V, is presented in Fig. 5. The response is measured as a relative response. As expected the sensitivity varies strongly with doping giving the lowest response for sample N3. The low response for N1 can be explained by the fact that the epitaxial layer in this wafer is very thin reducing the active depth. The response for the p-type sample is higher than for the corresponding n-type samples. The reason for this is that the diffusion length for electrons, which generate the photocurrent in the p-type sample, is larger than for holes and that the built in voltage in the p-type diode is much higher than in the n-type diode.

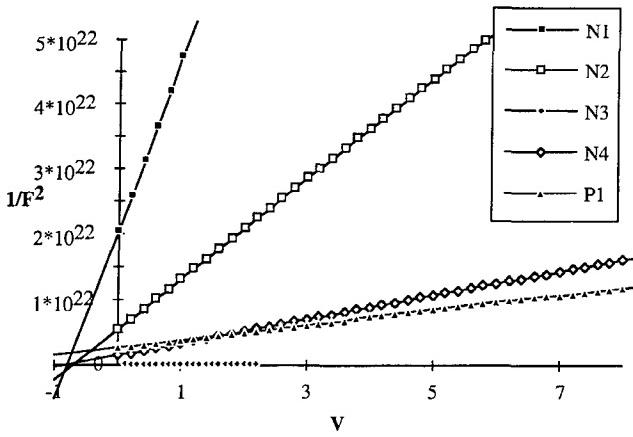


Fig. 3. CV-characteristics for all diodes.

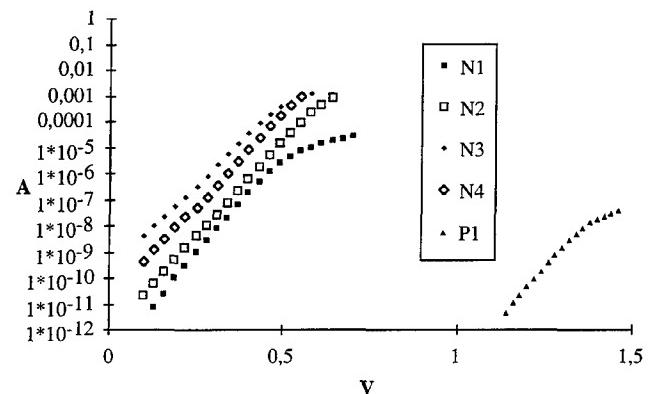


Fig. 4. IV-characteristics for all diodes.

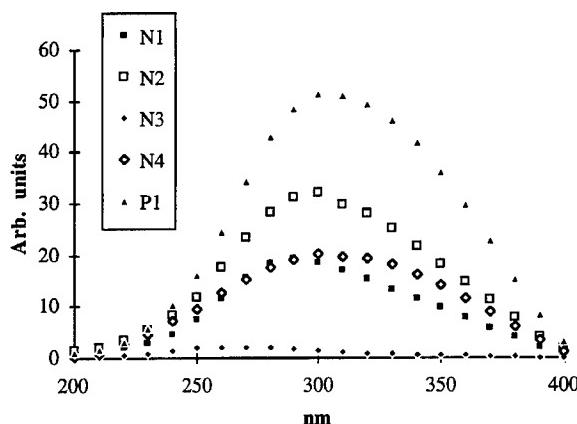


Fig. 5. Spectral sensitivity of the diodes. The diode on p-type material has the highest sensitivity and the diode on high doped n-type has the lowest sensitivity.

In Fig. 6 the response curves have been scaled to a peak response for each diode of 100. Thus the relative response for the different diodes can be compared. N2 has its peak sensitivity at 250 nm and shows very low sensitivity at longer wavelengths. This is also the sample where the shortest diffusion length could be expected. Sample N1 also shows a somewhat lower response at long wavelengths. In this case it can be explained by the shallow epitaxial layer.

5. Simulations and discussion

In order to investigate the theoretical sensitivity one segment of two different diodes, one on n-type and one on p-type material, was simulated using the program MEDICI. The simulated diode had a metal finger width of 1 μm and an opening between the fingers of 5 μm . The doping concen-

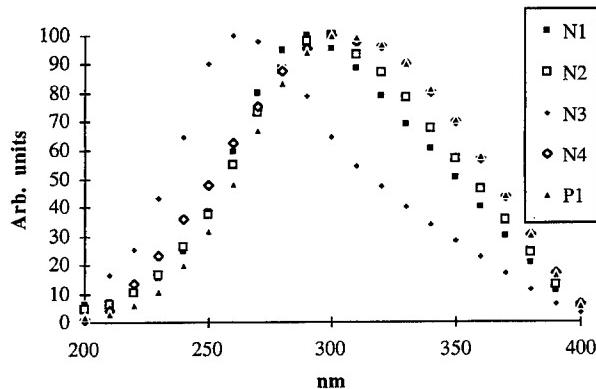


Fig. 6. Relative spectral sensitivity for all diodes. The highest measured response for each diode is taken as the value 100 for that diode. Peak sensitivity can be observed from this figure.

Table II. Simulated quantum efficiency for a n-type and a p-type diode. The table also includes the ratio of the response for the p-type diode versus the response for the n-type diode and the measured ratio for two of our diodes. The measured diodes have significantly higher doping than the simulated

	220 nm	300 nm	390 nm
N-type	0.8%	13.9%	1.8%
P-type	5.8%	63.6%	4.6%
P/N-sim	7.0	4.4	2.6
P/N-meas	2.6	2.5	1.2

tration was chosen to be $N = 1 * 10^{14} \text{ cm}^{-3}$ for both diodes and the carrier lifetime was set to 100 ns. These parameters were taken according to the previous simulations in [6]. The radiation is perpendicular to the surface of the semiconductor and the metal is considered opaque to the radiation. All generation of photocarriers then takes place in the semiconductor below the opening in the metal. The intensity of the radiation in this simulation was in the range of $10^{15} \text{ photons/cm}^2/\text{s}$. The surface recombination velocity was set to $1 * 10^7 \text{ cm/s}$. Simulation was done for the wavelengths 390, 300 and 220 nm. Short circuit currents were calculated. The absorption data used in our simulations were taken from [10].

The results from the simulations are presented in Table II. Results are presented as percent of incoming radiation contributing to the photocurrent. Reflection at the semiconductor surface is neglected. Due to the fact that one part of the semiconductor is shielded by the metal the maximum theoretical photoresponse is 83%.

The p-type material shows a much higher photoresponse than the n-type material. This can be explained by two effects. First, the built in voltage in the p-type material is much higher (about a factor of two) than the built in voltage in the n-type material. Second, the diffusion length for electrons is much larger than the diffusion length for holes giving a larger contribution from diffusion in the p-type material. This effect also explains why the difference in photocurrent is larger at long wavelengths where a large number of photons are absorbed deep down in the semiconductor. The reason for the low response at short wavelengths is the surface recombination.

If we compare the measurement results from a p-type diode, P1, with the result for a n-type diode with similar doping, N4, the we can see a similar effect even if it is less pronounced. The higher doping in these diodes will reduce the diffusion length.

6. Conclusions

Schottky diodes on 6H-SiC can be used as efficient photo detectors. The sensitivity of the diodes depends on doping concentration. p-type material gives higher short circuit currents due to larger built in voltages and larger diffusion lengths. Schottky diodes on 6H-SiC can be fabricated on substrates with as high doping as $2 * 10^{18} \text{ cm}^{-3}$ and still be used at reverse voltages of up to 10 V.

7. References

- Ivanov, P. A. and Chelnokov, V. E., Semicond. Sci. Technol. 7, 863 (1992).
- Davis, R. F. et al., Proceedings of IEEE 79, 677 (1991).
- Glasow, P. et al., SPIE 868, 40 (1987).
- Fröjd, C. et al., Presented at the 15th Nordic Semiconductor meeting, 1992 (unpublished).
- Brown, D. M. et al., IEEE Trans on Electron Devices Vol. 40, No. 2 February 1993.
- Fröjd, C., Thungström, G., Nilsson, H-E. and Petersson, C. S., MRS Proceedings Vol. 339, 1994.
- Bhatnagar, et al., IEDM 92, 789 (1992).
- Lundberg, N. and Östling, M., Appl. Phys. Lett. 63, 3069 (1993).
- Lundberg, N., Zetterling, C. M. and Östling, M., Appl. Surf. Sci. 73, 316 (1993).
- Choyke, W. C. and Palik, E. E., "Silicon Carbide (SiC) in Handbook of Optical Constants of Solids" (Academic press, 1985), pp. 587-595.

Large Area GaInAsP and GaInP Solar Cells for Space Applications

K. Smekalin, K. Tappura and J. Lammastniemi

Tampere University of Technology, Department of Physics, P.O. Box 692, FIN-33101 Tampere, Finland

Received May 4, 1994; accepted June 15, 1994

Abstract

We report the fabrication and characterization of large area ($1 \times 1 \text{ cm}^2$) $\text{Ga}_{0.84}\text{In}_{0.16}\text{As}_{0.68}\text{P}_{0.32}$ ($E_g = 1.50 \text{ eV}$) and $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$ ($E_g = 1.88 \text{ eV}$) solar cells. The cell structures were grown by gas-source MBE on 2" (100) GaAs substrates. For the GaInAsP material, both n-on-p and p-on-n structures having wide-gap $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$ window were studied. The GaInAsP n-on-p cells showed significantly better active area conversion efficiencies (17.8% at AM0, 1-sun illumination) than p-on-n structures (13.0%, same conditions) mostly due to lower sheet resistance of the n-type GaInAsP emitter layers. For GaInP cells the best conversion efficiency of 11.0% was achieved for windowless shallow homojunction n-on-p structure. Since only single layer of SiN_x was utilized as an antireflection coating for all the cells, we believe that the application of an optimized two-layer antireflection coatings could increase the efficiencies up to 19% and 14% for GaInAsP and GaInP solar cells, correspondingly. The excellent uniformity in all the cell parameters across the 2" wafers indicates that larger area solar cells (up to 10 cm^2) can be fabricated.

1. Introduction

Phosphorus containing III-V semiconductors, such as InP [1], GaInP [2] and GaInAsP [3], have shown excellent stability of their material parameters when irradiated by high-energy particles and therefore they are considered as good candidates for space solar cell materials. In comparison to InP, GaInAsP and GaInP materials have the advantage that they can be grown lattice-matched on GaAs or Ge substrates. These substrate materials have better mechanical strength and they can be made thinner and therefore lighter than InP substrates.

$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$ material has wide opportunities to vary the band-gap and lattice constant. Quaternary material with the composition of $x = 0.84$ and $y = 0.68$ has nearly optimum energy band gap (1.5 eV) for conversion of AM0 sunlight illumination [4]. It can also be used for making monolithic tandem solar cell structures, because it is both current-matched and lattice-matched to Ge [3]. Such a tandem structure is expected to have efficiency around 30% [3, 4].

Lattice matched $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$ having a band-gap of 1.9 eV can be used as a top cell in a monolithic tandem solar cell structure with bottom cell made of GaAs, because these cells can be adjusted to be current-matched under AM0 illumination. Such tandem structure has already been reported with efficiency over 27% at AM1.5 [5].

In this paper we report the fabrication and characterization of large area ($1 \times 1 \text{ cm}^2$) $\text{Ga}_{0.84}\text{In}_{0.16}\text{As}_{0.68}\text{P}_{0.32}$ and $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$ solar cells on GaAs substrates with the

peak efficiency of 17.8% for GaInAsP cells (AM0, one sun, active area) and 11.0% for GaInP cells (same illumination). The uniformity of the average efficiency of the cells across the whole 2" wafer points out that larger area solar cells can be manufactured. The possibilities for further cell efficiency improvement are discussed.

2. Experimental

A schematic of the solar cell structures on GaAs substrates is shown in Fig. 1. The structures have much in common; the main difference is that the $\text{Ga}_{0.84}\text{In}_{0.16}\text{As}_{0.68}\text{P}_{0.32}$ cell [Fig. 1(a)] has a thin window on the top and a back surface field (BSF) layer both made of wide-gap $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$. The lattice-matched GaInP window is known to improve spectral response in blue region for both GaAs [6] and GaInAsP cells. The $\text{Ga}_{0.51}\text{In}_{0.49}$ structure [Fig. 1(b)], on the contrary, has a shallow p-n junction without any window layer and the back field barrier is formed by a heavily doped GaInP BSF layer. Although the AlInP wide-gap window and BSF layers have been recently reported for GaInP cells [7, 8], poor quality AlInP/GaInP interface can hinder the improvement of such cells [7]. Therefore the shallow homojunction structure for the GaInP cell was chosen.

The solar cell structures were grown by gas-source molecular beam epitaxy (GSMBE) using the dual reactor chamber V80H machine built by VG Semicon. The diffusion pumped GSMBE chamber is equipped with a high pressure cracking

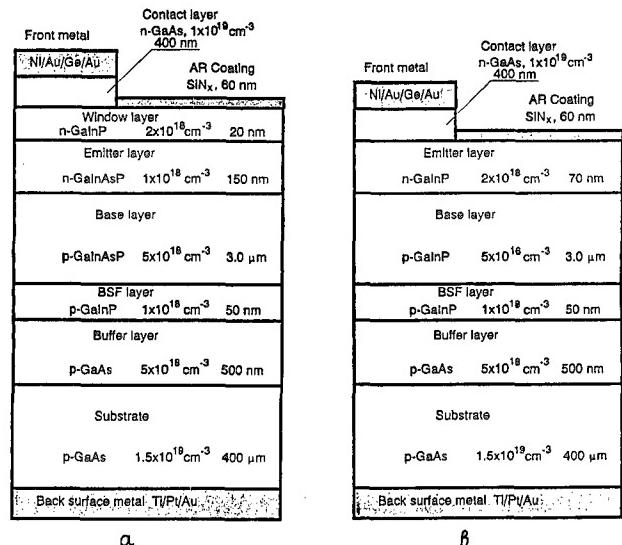


Fig. 1. The GaInAsP (a) and GaInP (b) solar cell structures. Similar structure as (a) was utilized for p-on-n GaInAsP solar cells.

cell (at 950 °C) for thermal decomposing of the group V hydrides, AsH_3 and PH_3 , used as group V source materials. Solid source effusion cells are used to generate the group III and dopant material beams. There are two effusion cells for indium, one for gallium and one for both of the dopant sources, silicon and beryllium. In and Ga fluxes were measured with a calibrated ion gauge moved to the substrate position for the measurement. The structures were grown on 2-inch GaAs(001) wafers attached to the substrate holder without using indium solder. The "epi-ready" GaAs wafers were loaded in vacuum without any wet-chemical pretreatment. A pre-growth heat treatment was performed in the vacuum chamber at a substrate temperature of 600–610 °C under arsenic pressure. A narrow-band infrared pyrometer was used to measure the substrate temperature, which was between 500 and 510 °C during the growth of GaInAsP and GaInP layers and about 600 °C for GaAs layers. The growth rates were about 1.4, 1.7 and 2.8 $\mu\text{m}/\text{h}$ for GaAs, GaInAsP and GaInP, respectively. Rotation was performed for the samples during growth in order to obtain a homogeneous distribution of group III materials over the entire sample area.

The lattice matching was checked by measuring X-ray diffraction rocking curves (Fig. 2). The curves reveal reasonably good crystalline quality for both GaInAsP and GaInP structures. A small compressive strain of approximately 0.02% at room temperature, providing the lattice matching at growth temperature, was considered as an optimum. The band-gap energy of the quaternary material was evaluated by measuring room-temperature photoluminescence (PL) spectrum of a single GaInAsP and GaInP test layers grown prior to the actual structures at the same growth conditions.

Contacts to n-type GaAs were made by sequential evaporation of Ni/Au/Ge/Au metal composition. The contact was annealed at 400 °C for 1 minute. For p-type contacts, Ti/Pt/Au layers were evaporated. After evaporating and annealing both contacts were electroplated with gold up to the thickness of approximately 2 μm . The front contact grid was defined by photolithography with an obscuration of 20% of the total cell area. The GaAs contact layer was removed from between the metal grid lines by selective etching the GaAs over the GaInP layer with phosphoric acid : hydrogen peroxide : water (1 : 1 : 25) solution. The cell perimeter was also defined by photolithography and by

mesa etching the GaInP layer with hydrochloric acid : phosphoric acid (1 : 1) solution, and for the quaternary structure the GaInAsP material with sulphuric acid : peroxide : water (1 : 1 : 1) solution. A single layer of SiN_x with a thickness of 60 nm was applied to the front surface of the device to form the antireflection coating and to protect mesa edges. The total area of the device is 1 cm^2 . Four 1 \times 1 cm^2 solar cells were normally fabricated out of one 2" wafer, along with a number of smaller cells.

The power conversion efficiencies of the solar cells were measured under approximate AM0 illumination. Some of the data obtained were also confirmed at EEV Ltd. using an AM0 simulator. Since the cells have relatively large front grid coverage, all the experimental data on spectral responses and power conversion efficiencies were corrected according to grid loss and refer hereafter to the active area of the devices.

3. Results and discussion

Figure 3 represents the data on spectral response for both GaInAsP and GaInP cells. As it can be seen, the GaInP shallow homojunction cell [Fig. 3(a)] has reasonably good response in the blue region, due to low absorption of light in the thin emitter layer. For the GaInAsP cells it was observed that n-on-p structure [Fig. 3(b)] has better quantum efficiency than the p-on-n one for all wavelengths. The main reason for the lower response of the p-on-n structure is the high sheet resistance of the p-type GaInAsP emitter layer. It was proved by direct measurement of the sheet resistance of both layers using the conventional transient line method. For the n-type GaInAsP emitter layer the specific resistance appeared to be $3 \times 10^{-3} \Omega \text{ cm}$, while for the p-type it was $0.1 \Omega \text{ m}$. Using the known doping levels for both layers, we are able to derive an estimation for the free carrier mobilities in these materials. For n-type electrons it is around $1200 \text{ cm}^2/\text{V sec}$, and for p-type holes it was only $40 \text{ cm}^2/\text{V sec}$. These values are in a good agreement with Hall mobilities of the single GaInAsP epitaxial layers grown by GSMBE. Therefore p-type GaInAsP appears to be inappropriate material for thin emitter layers. The problem of the emitter sheet resistance cannot also be solved by making the emitter layer thicker, because both experiment and cal-

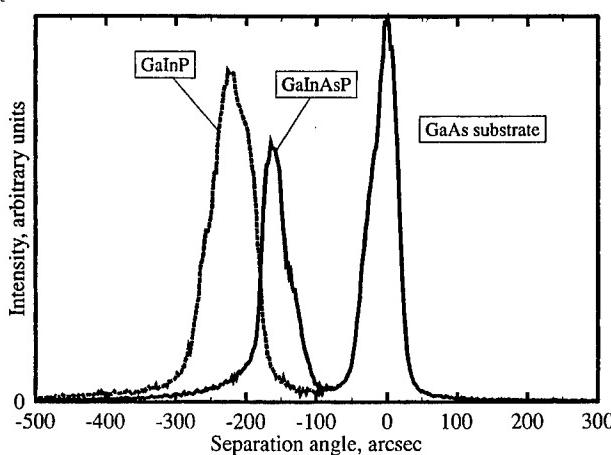


Fig. 2. X-ray diffraction rocking curves for GaInAsP (a) and GaInP (b) solar cell structures.

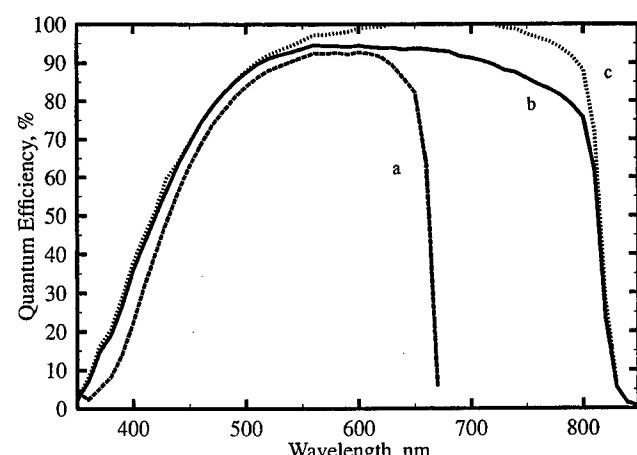


Fig. 3. Quantum efficiencies (QEs) of the solar cells: a – external QE of the GaInP cell; b – external QE of the GaInAsP cell, c – internal QE for the GaInAsP cell.

culation show that thicker emitter layers will result in additional decrease in the cell efficiency.

Gradual decrease in quantum efficiency that appears on the long-wavelength side QE spectrum of the GaInAsP cell [Fig. 3(b)] is due to the reflection from the surface of the structure. The thickness of the single layer SiN_x anti-reflection coating was chosen to minimize the reflection at around 500 nm, therefore the reflection for wavelength over 680 nm exceeds 10%. This can be seen from the internal quantum response spectrum for the same structure [Fig. 3(c)]. Changing the antireflection coating to two-layer system, e.g. ZnS/MgF_2 , can sufficiently improve the spectral response and thus increase the total conversion efficiency. The decrease in response of the GaInP cell for the wavelength over 600 nm [Fig. 3(a)] is attributed to relatively short diffusion length of minority electrons in GaInP base layer. It is reported to be about 1–1.5 μm [8] at best. Therefore, when part of the incident light is absorbed deeply in the base layer, as it happens for the wavelengths over 600 nm, not all the photogenerated electrons can reach the p-n junction and contribute to the photocurrent.

The illuminated current-voltage (*IV*) characteristics of both the GaInAsP and GaInP solar cells under AM0 spectrum, 1-sun intensity (135 mW/cm^2) is shown in Fig. 4. The open-circuit voltage U_{oc} , fill-factor (FF) and active area conversion efficiency were 1.014 V, 0.87 and 17.8% for n-on-p GaInAsP device, and 1.200 V, 0.80 and 11.0% for GaInP structure, respectively. The open-circuit voltage of 1.2 V appears to be somewhat low in comparison to the values of up to 1.3 V for MOCVD-grown structures reported recently [2, 8]. This might be due to the certain leakage across the p-n junction caused by material imperfection of the gas-source MBE grown GaInP. We believe two main reasons are responsible for that. First, the incorporation of the p-dopant (Be) in GaInP layers that can cause clusters formation and distortion of the crystal lattice. Second, the ordering/disordering of the GaInP alloy that can change across the structure during the growth. The GaInP material study is under way now, and we believe that further

improvement of the GaInP material properties should make it possible to increase the open-circuit voltage and the efficiency of the GaInP cells.

Having obtained the experimental data on internal quantum efficiencies for both cells, we estimate that an appropriate two-layer antireflection coating for the devices, e.g. ZnS/MgF_2 could increase the active area conversion efficiency up to 19% for GaInAsP and 14% for GaInP. Having processed and studied a number of 2" wafers into solar cell devices, we also made a conclusion that for all the cells processed from the same wafer the efficiency figures lay within no more than 0.3% deviation from the average figure for the wafer. This is a clear indication that even larger area solar cells (up to 10 cm^2) can be fabricated out of a 2" wafer grown by gas-source MBE.

4. Conclusion

In summary, we have reported the fabrication and characterization of large area ($1 \times 1 \text{ cm}^2$) $\text{Ga}_{0.84}\text{In}_{0.16}\text{As}_{0.68}\text{P}_{0.32}$ and $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$ solar cells on GaAs substrates. The best efficiency achieved for the structures having single-layer SiN_x antireflection coating was 17.8% for GaInAsP and 11.0% for InGaP at AM0, 1 sun illumination. The application of an optimized two-layer antireflection coating should increase the efficiency of the device up to 19% and 14%, respectively. Also the excellent uniformity in cell efficiencies across the 2" wafer makes it possible to fabricate larger area (up to $3.5 \times 3.5 \text{ cm}^2$) solar cells. The data obtained shows that gas-source MBE grown GaInAsP and GaInP materials are competitive candidates for the next generations of space solar energy systems.

Acknowledgements

The authors wish to thank Dr. T. Cross and Mr. C. Hardingham at EEV Ltd. (England) for their co-operation and confirmation of the results obtained in this work. This work was partially supported by EEV Ltd.

References

- Yamamoto, A., Yamaguchi, M. and Uemura, C., *Appl. Phys. Lett.* **44**, 611 (1984).
- Chiang, P. K., Vijayakumar, P. S. and Cavicchi, B. T., "23rd IEEE Photovoltaic Specialists Conference" (IEEE, New York 1993), p. 659.
- Sharps, P. R. et al., "23rd IEEE Photovoltaic Specialists Conference" (IEEE, New York 1993), p. 633.
- Jain, R. K. and Flood, D. J., *J. Solar Energy Engineering* **115**, 106 (1993).
- Olson, J. M., Kurtz, S. R., Kibbler A. E. and Faine, P., *Appl. Phys. Lett.* **56**, 623 (1990).
- Olson, J. M., Ahrenkiel, R. K., Dunlavy, D. J., Keyes, Brian and Kibbler A. E., *Appl. Phys. Lett.* **55**, 1208 (1989).
- Wojtczuk, S. J., Vernon, S. M. and Sanfacon M. M., "23rd IEEE Photovoltaic Specialists Conference" (IEEE, New York 1993), p. 655.
- Ikeda, E., Takamoto, T., Kurita, H. and Ohmori, M., "6th International Conference on Indium Phosphide and Related Materials" (IEEE, New York 1994), p. 500.

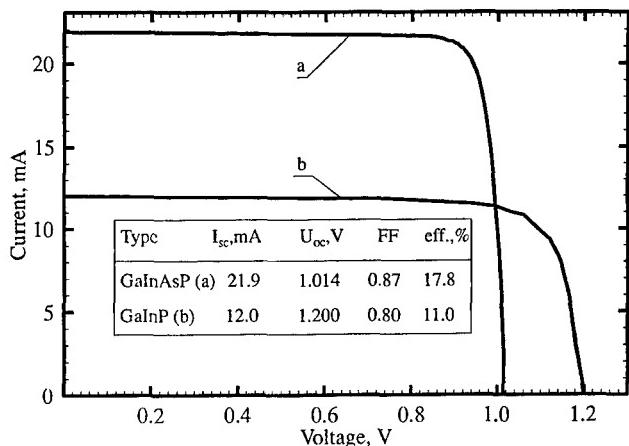


Fig. 4. IV-characteristics of the cells under approximate AM0, 1 sun illumination: a – GaInAsP cell; b – GaInP cell. The inset shows cell parameters for both configurations.

Fitting of the Solar Cell /V-curve to the Two Diode Model

A. Hovinen

Helsinki University of Technology, Electron Physics Laboratory, Otakaari 7 A, SF-02150, Espoo, Finland

Received May 5, 1994; accepted June 15, 1994

Abstract

In this paper, the parameters for the solar cell two diode model are solved with respect to the cell series resistance, thus providing an efficient mean to extract the possible combinations for the model parameters. In the calculations, the diode ideality factors n_1 and n_2 have been left fixed. The usual way is to set $n_1 = 1$ and $n_2 = 2$ to represent the diffusion and recombination current terms, respectively. The procedure is tested with respect to some published cell parameter values with good results.

1. Introduction

The current versus voltage characteristics of a solar cell under illumination is commonly translated to an equivalent circuit containing a photocurrent source and a diode with a shunt resistor, and a series resistor in the load branch. To get a more accurate fit to a measurement data, a second parallel diode is sometimes added to the circuit. The two diode model allows one to separate the diffusion current term, in which the diode ideality factor n equals 1, and the recombination current term, in which the ideality factor equals 2. The other model parameters are the photocurrent, I_{ph} , the diode saturation currents I_{s1} and I_{s2} , the shunt resistance R_{sh} , and the cell series resistance R_s .

The task for extraction of the five model parameters from the cell current equation is not analytically solvable, and iterative methods must be used to fit the measured current versus voltage curve to the equation. The fit can be based on a set of five equations derived from the current equation in different ways. A common choice is to fit the parameters in the open circuit point ($0, V_{oc}$), short circuit point ($I_{sc}, 0$) and maximum power point (I_m, V_m), with the remaining two equations derived from the curve slope at ($0, V_{oc}$) and ($I_{sc}, 0$). One way to proceed to the solution is to solve as many parameters as possible with respect to a smaller set of parameters and iterate from this limited set of equations. With a proper choice of starting equations, four of the parameters can be presented as a function of the series resistance only, as presented in this paper. The series resistance R_s remains to be iterated. As the parameters I_{ph} , I_{s1} , I_{s2} and R_{sh} are represented as functions of R_s only, and should be positive in sign in the correct solution, one has a very effective mean to narrow out the R_s ranges inside which the solutions are to be found.

2. Equations used in the parameter extraction procedure

The current equation for the two diode model $F(I, V)$ is

$$F(I, V) = I - I_{ph} + I_{s1} \cdot (e^{(V+I \cdot R_s)/n_1 \cdot V_t} - 1) + I_{s2} \times (e^{(V+I \cdot R_s)/n_2 \cdot V_t} - 1) + \frac{V + I \cdot R_s}{R_{sh}} = 0,$$

where I is the cell output current, I_{ph} is the photocurrent, I_{s1} , I_{s2} are the saturation currents of diodes 1 and 2, R_s the series resistance, R_{sh} the shunt resistance, V is the output voltage of the cell, and $V_t = kT/q$ where k is the Boltzmann constant, T the cell temperature and q the electron charge. In this paper the diode ideality factors n_1 and n_2 are fixed to 1 and 2 to represent the diffusion and recombination current terms, respectively [3].

The five base equations are now derived in a somewhat similar manner as was done by Chan *et al.* [1] with the one diode model, and Enebishi *et al.* [2] with the two diode model. Three of the five equations can be found putting the pairs $(0, V_{oc})$, $(I_{sc}, 0)$, (I_m, V_m) to $F(I, V)$. In the derivation of the remaining two formulas, one can use the relation [2]

$$\begin{aligned} \frac{dF}{dI} + \frac{dF}{dV} \frac{dV}{dI} \\ = 1 + \left(R_s + \frac{dV}{dI} \right) \\ \times \left(\frac{I_{s1}}{n_1 V_t} e^{(V+IR_s)/n_1 V_t} + \frac{I_{s2}}{n_2 V_t} e^{(V+IR_s)/n_2 V_t} + \frac{1}{R_{sh}} \right) = 0 \end{aligned}$$

and solve it in the short circuit ($I_{sc}, 0$) and maximum power (I_m, V_m) points. To shorten the expressions, we now adopt the following notation:

$$A = e^{V_{oc}/n_1 \cdot V_t}, \quad B = e^{V_{oc}/n_2 \cdot V_t}, \quad C = e^{I_{sc} \cdot R_s/n_1 \cdot V_t} \\ D = e^{I_{sc} \cdot R_s/n_2 \cdot V_t}, \quad E = e^{(V_m + I_m \cdot R_s)/n_1 \cdot V_t}, \quad G = e^{(V_m + I_m \cdot R_s)/n_2 \cdot V_t}.$$

We also write

$$\begin{aligned} \alpha &= C \cdot \left(1 + \frac{V_{oc} - I_{sc} \cdot R_s}{n_1 \cdot V_t} \right) - A, \\ \beta &= D \cdot \left(1 + \frac{V_{oc} - I_{sc} \cdot R_s}{n_2 \cdot V_t} \right) - B \\ \gamma &= E - A - C \cdot \frac{V_m + I_m \cdot R_s - V_{oc}}{n_1 \cdot V_t}, \\ \delta &= G - B - D \cdot \frac{V_m + I_m \cdot R_s - V_{oc}}{n_2 \cdot V_t} \end{aligned}$$

Note that the expressions are functions of R_s only. Next, we define $dV/dI_{V=0} \equiv R_{sh0}$ to represent the diode shunt resistance as determined directly from the IV -curve slope in the short circuit ($I_{sc}, 0$) point. After some mechanical calculations we can represent I_{s1} , I_{s2} , I_{ph} , and R_{sh} as a function of R_s only:

$$I_{s1} = \frac{\delta I_{sc} - 1/(R_{sh0} - R_s) (\beta(V_m + I_m R_s - V_{oc}) + \delta(V_{oc} - I_{sc} R_s)) - \beta I_m}{\gamma \cdot \beta - \alpha \cdot \delta}$$

$$I_{s2} = \frac{1}{\beta} \cdot \left(\frac{V_{oc} - I_{sc} \cdot R_s}{R_{sh0} - R_s} - I_{sc} - \alpha \cdot I_{s1} \right)$$

$$I_{ph} = I_{s1} \left(A - 1 - \frac{V_{oc}}{n_1 \cdot V_t} \cdot C \right) + I_{s2}$$

$$\times \left(B - 1 - \frac{V_{oc}}{n_2 \cdot V_t} \cdot D \right) + \frac{V_{oc}}{R_{sh0} - R_s}$$

$$\frac{1}{R_{sh}} = \frac{1}{R_{sh0} - R_s} - I_{s1} \cdot \frac{C}{n_1 \cdot V_t} - I_{s2} \cdot \frac{D}{n_2 \cdot V_t}.$$

The R_s can be iterated from the formula

$$1 + \left(R_s - \frac{V_M}{I_M} \right) \cdot \left(\frac{I_{s1}}{n_1 \cdot V_t} \cdot E + \frac{I_{s2}}{n_2 \cdot V_t} \cdot G + \frac{1}{R_{sh}} \right) = 0.$$

3. Results and discussion

A procedure to make the parameter extraction has been written in C. The input parameters defining the specific cell under study are the short circuit current I_{sc} , the open circuit voltage V_{oc} , the voltage V_M and current I_M in the maximum power point, the voltage versus current curve slope at short circuit R_{sh0} , and temperature T . The parameter extraction is initiated by checking the ranges of R_s inside where positive values of I_{ph} , I_{s1} , I_{s2} and R_{sh} can be found simultaneously. The possible solutions for R_s is then iterated inside these ranges.

The procedure was tested against published parameter values of a pn-junction solar cell. P. Sana, J. Salami and A.

Rogathi [4] have fabricated high efficiency polycrystalline silicon solar cells and used NREL to make the standard measurements. Their NREL measurements at 25°C on a 1.00 cm² 17.7% efficient cell gave $V_{oc} = 0.6259$ V, $I_{sc} = 35.60$ mA, $V_M = 0.5284$ V, $I_M = 33.4$ mA. The shunt and series resistance measurements gave $R_{sh} = 299$ kohm and $R_s = 0.34$ ohm.

The group also analyzed the results with a two diode model using ideality factor of 3.6 with the second diode. Their calculations gave $I_{s1} = 1.33 \cdot 10^{-12}$ A and $I_{s2} = 9.37 \cdot 10^{-7}$ A for the diode saturation currents. The corresponding values from our procedure were $I_{s1} = 8.77 \cdot 10^{-13}$ A and $I_{s2} = 6.77 \cdot 10^{-7}$ A, which are very close to the reference, despite the fact that $n_1 = 1$ had to be assumed as in the reference paper its value was not mentioned. The R_s value obtained was 0.487 ohm, which is relatively far away from the 0.34 ohm measured reference value. In practical use, one of the most critical things with the procedure is the accurateness of the measured input parameters, e.g. T and R_{sh0} , as was pointed out in [1].

References

1. Chan, D. S. H., Phillips, J. R. and Phang, J. C. H., Solid State Electron. **29**, 329 (1986).
2. Enebishi *et al.*, Solar Energy Materials and Solar Cells **29**, 201 (1993).
3. See for instance S. M. Sze, Physics of Semiconductor Devices, 2nd ed., New York, 1981, John Wiley & Sons.
4. Sana, P., Salami, J. and Rohatgi, A., IEEE Trans. Electron Dev. **40**, 1461 (1993).

Development of Materials for Blue/Green Light Emitters

K. Rakennus, P. Uusimaa, K. Smekalin, P. Savolainen and M. Pessa

Department of Physics, Tampere University of Technology, P.O. Box 692, FIN-33101 Tampere, Finland

Received May 4, 1994; accepted June 15, 1994

Abstract

We have adopted a new approach in an attempt to improve the present-day technology in preparing blue/green ZnSe light emitters. This approach includes growth of layer structure with n-on-p configuration and the use of barrier reduction layers in between the GaAs substrate and the lower cladding layer of the device structure. In addition, we have studied MnZnSSe quaternary alloys to replace the MgZnSSe layers which are normally used as cladding layers of the blue/green light emitters and which are known to exhibit poor electrical properties.

1. Introduction

The use of ZnSe and related II–VI compound semiconductors, such as ZnSSe and CdZnSe grown by molecular beam epitaxy (MBE), together with a nitrogen rf plasma source [1], has resulted in blue/green quantum well (QW) lasers [2, 3] and light emitting diodes (LEDs).

Preparing the materials for the II–VI lasers and LED's is very difficult. One of the difficulties is to obtain the lattice matching of II–VI layers to each other and to the substrate crystal structure. For example, the critical thickness of an MBE-grown ZnSe layer on a GaAs substrate having the lattice mismatch of 0.27% is about 1500 Å [4]. As soon as the ZnSe film becomes thicker than the critical thickness it relaxes, and a high density of misfit dislocations is resulted in. Thick epitaxial layers, much thicker than 1500 Å, which were free of dislocations are necessary for making devices. Therefore, growth of such layers is only possible by using ternary (ZnSSe) or quaternary (MgZnSSe) alloys which can be lattice matched to GaAs. For example, separate confinement heterostructure (SCH) blue/green quantum well laser diodes usually consist of lattice matched $Mg_{0.09}Zn_{0.91}S_{0.12}Se_{0.88}$ quaternaries as cladding layers and lattice matched $ZnS_{0.07}Se_{0.93}$ as waveguides [5].

Typically, ZnSe nucleates in a three-dimensional (3-D) manner in the initial stage of layer growth on GaAs. However, depositing ZnSe onto a GaAs buffer layer, which is grown on a GaAs substrate, 2-D nucleation can be achieved at a very early stage of growth [6].

Because MgZnSSe used as a cladding layer of laser or LED is poor in electrical quality we have studied MnZnSSe as a replacement for MgZnSSe. In addition, we have attempted to overcome the difficulties encountered in the conventional p-on-n layer configuration by preparing n-on-p structures since no ohmic contact to p-type ZnSe is possible. On the other hand, in the n-on-p configuration there exists a large potential barrier for holes at the p-GaAs/p-ZnSe interface. Our aim is to reduce this valence band offset

by growing III–V barrier reduction layers, which open the band gap (E_g) to make it closer to E_g of ZnSe.

In the present paper we discuss structural and optical properties of ZnSe-based materials grown on GaAs substrates by MBE.

2. Experimental

ZnSe-based layers and structures were grown in a DCA Instruments MBE system equipped with the reflection high-energy electron diffraction (RHEED) technique. Elemental Zn, Se, Cd and Mn and the ZnS compound were used as source materials. A $ZnCl_2$ source was used for n-type doping, while p-type doping was achieved using an Oxford Applied Research rf nitrogen plasma source. The beam fluxes were measured with a nude ion gauge. The fluxes were set to give a beam equivalent pressure ratio $Zn : Se \approx 1.3$ for growth of ZnSe. Optimum (nominal) growth temperature of ZnSe was found to be around 350 °C, and of ZnSSe and other compounds about 300 °C. The GaAs (100) substrates used in these experiments were deliberately misoriented 2° off towards (110).

GaAs buffer layers and III–V barrier reduction layers were grown in a double-chamber MBE system VG from Semicon Ltd. One of the VG chambers comprised a gas-source MBE (GSMBE), the other chamber comprised a conventional solid source MBE. The III–V layers were grown on GaAs by the GSMBE, and were then transferred in vacuum into the MBE chamber for deposition of an arsenic cap layer. The As cap layer protected the surface of the sample during the sample transfer in air from this reactor to the II–VI reactor.

After growth of the II–VI layers, the total film thickness was measured with a stylus surface profiler. Composition and crystalline quality were studied by double-crystal X-ray diffraction (DCXRD). Optical and electrical properties were studied using photoluminescence (PL), Hall measurements, and electrochemical capacitance-voltage (ECV) profiling.

3. Results

3.1. ZnSE

A number of relatively thick ZnSe films (1 μm–4 μm) were grown to determine the effects of various growth parameters on crystallinity of ZnSe and to study the electrical properties of ZnSe : Cl. All the layers grown, including ternary and quaternary alloys, possessed shiny mirror-like surfaces.

A spotty "fishnet" RHEED pattern observed in the initial stage of growth the ZnSe suggested an occurrence of three dimensional nucleation when depositing the layers directly on GaAs substrates. On the contrary, growth of ZnSe on GaAs buffer layers, prepared as described above, resulted in the formation of streaky (2×1) and c(2×2) reconstructed RHEED patterns at very early stage of growth indicating a 2-D character of nucleation.

When the ZnSe layer was relaxed (thickness over 2000 Å) the full width at half maximum (FWHM) of the $\{400\}$ diffraction peak of DCXRD was very broad, due to the presence of misfit dislocations. However, the FWHM became narrower, as the layer thickness was increased. The FWHM decreased from 300 arc sec for a $2 \mu\text{m}$ ZnSe to 120 arc sec for a $4 \mu\text{m}$ ZnSe.

Figure 1 shows the carrier concentration (n) and the Hall mobility of ZnSe, determined at 300 K, as a function of ZnCl_2 cell temperature. The Hall mobility is about $400 \text{ cm}^2/(\text{Vs})$ at a doping level of $n = 4 \times 10^{16} \text{ cm}^{-3}$ and $230 \text{ cm}^2/(\text{Vs})$ at $n = 4 \times 10^{18} \text{ cm}^{-3}$ which was the highest doping level achieved.

3.2. ZnSSe

The mole fraction x in $\text{ZnS}_x\text{Se}_{1-x}$ must be about 7% to yield lattice matching to GaAs at growth temperature. Figure 2 shows a typical DCXRD rocking curve obtained for a $1.0 \mu\text{m}$ thick ZnSSe film on a GaAs buffer layer. The FWHM is 60 arc sec. For comparison, similar layers grown directly on GaAs substrates without the GaAs buffer layers resulted in much poorer crystal quality with the FWHM of more than 100 arc sec. This comparison points out that the initial stage of II-VI growth is critical.

Prior to ZnSSe growth, a thin (200 Å) ZnSe layer was always grown on GaAs in order to prevent the pitting caused by sulphur which attacks a bare GaAs surface in the growth chamber [6].

Nitrogen doping experiments were performed mainly for ZnSSe samples. The rf-power of the nitrogen source was varied between 120 and 350 W, while the nitrogen partial pressure was about 3×10^{-6} mbar. Since direct Hall mobil-

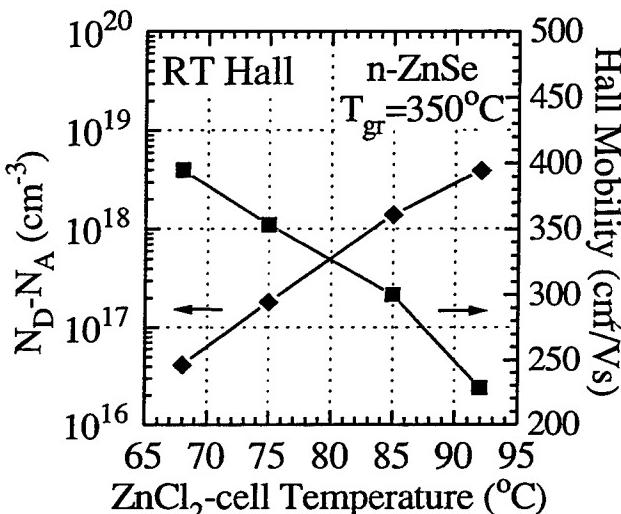


Fig. 1. Room temperature Hall measurements for Cl-doped ZnSe epitayers. Electrical properties of these n-type layers meet the requirements of practical device structures.

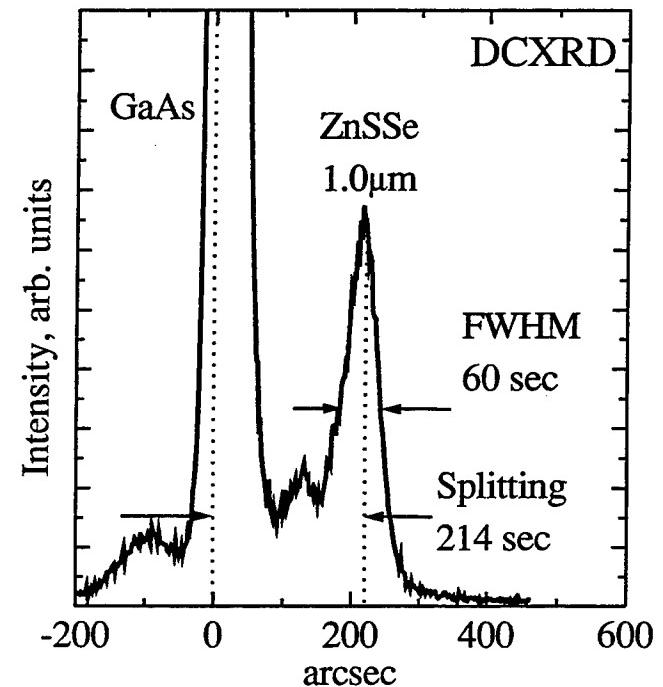


Fig. 2. DCXRD-curve for a $1 \mu\text{m}$ thick closely lattice-matched ZnSSe-layer grown on a GaAs-epilayer.

ity measurements were not possible for these samples, due to the problems with ohmic contacts, we applied an ECV technique to study the p-type conductivity. The ECV data for ZnSSe:N are shown in Fig. 3. As one can see, the largest net acceptor concentration was $1 \times 10^{17} \text{ cm}^{-3}$ obtained at an rf power of 200 W.

3.3. MnZnSSe

Our preliminary experiments on growth of $\text{Mn}_x\text{Zn}_{1-x}\text{S}_{1-y}$ have been quite promising. In these experiments we

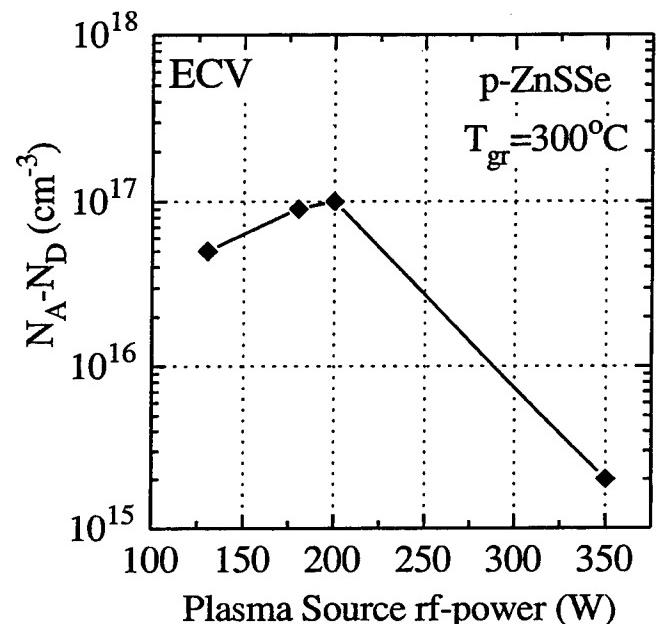


Fig. 3. Net acceptor concentration of p-ZnSSe measured by electrochemical CV-profiler. The net carrier concentration drops dramatically at rf powers over 200 W, indicating a rapid increase in compensation.

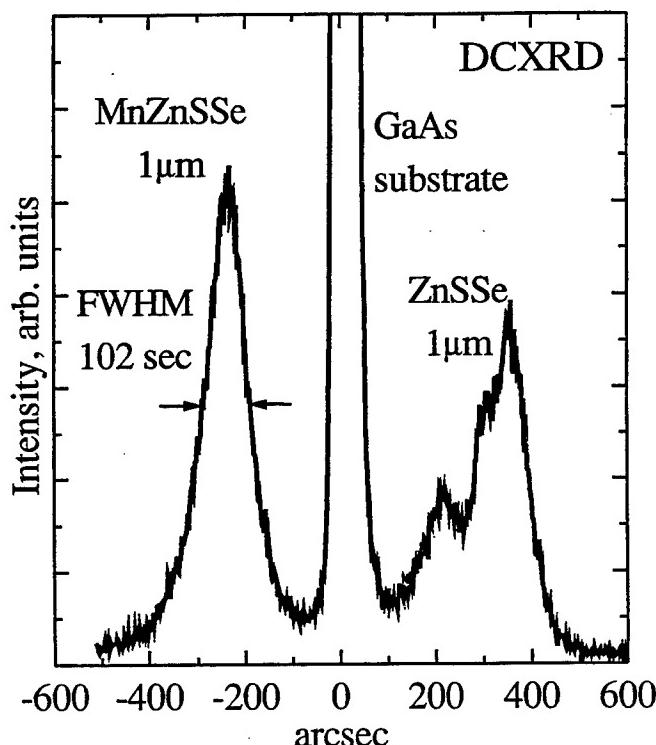


Fig. 4. DCXRD-curve for a MnZnSSe-layer grown on a ZnSSe/GaAs substrate.

first deposited a thin ZnSe protection cap layer on a GaAs buffer layer, and then grew 1 μm thick ZnSSe and 1 μm MnZnSSe layers. A DCXRD scan in Fig. 4 reveals that the structural quality of MnZnSSe is reasonably good. From this DCXRD it was concluded that x was about 0.05 and $y \approx 0.10$. From optical absorption studies, the band gap was determined to be about 2.9 eV.

3.4. CdZnSe quantum well

High-efficiency semiconductor light emitters requires quantum well structures. We have studied CdZnSe as a QW material. Our typical multi quantum well structure contained five 70 Å strained-layer Cd_{0.18}Zn_{0.82}Se QWs separated by 100 Å Zn_{0.07}Se_{0.93} barrier layers, all grown on a 1.5 μm thick Zn_{0.07}Se_{0.93} buffer layer.

Figure 5 shows PL spectra for a 5-QW ZnSSe/CdZnSe sample at 300 K and 80 K. The main peaks in both spectra are due to 1e-1hh transitions. The high-energy shoulder in the 300 K PL spectrum is assigned to 1e-1lh transitions. The low PL intensity of the ZnSSe barrier material seen in the 300 K PL shows that there is a strong quantum confinement of carriers in this QW-system.

3.5. Effects of barrier reduction layers

When a layer structure with n-on-p polarity is grown the large valence band offset of about 1 eV formed at the p-GaAs/p-ZnSe interface must be lowered in order to inject holes effectively into the QW region of a device. This potential barrier can be lowered by growing barrier reduction layers the energy bandgaps of which increase stepwise as one approaches the interface between the III-V and II-VI.

We have theoretically studied the effects of barrier reduction layers on the electrical properties of a pn-junction

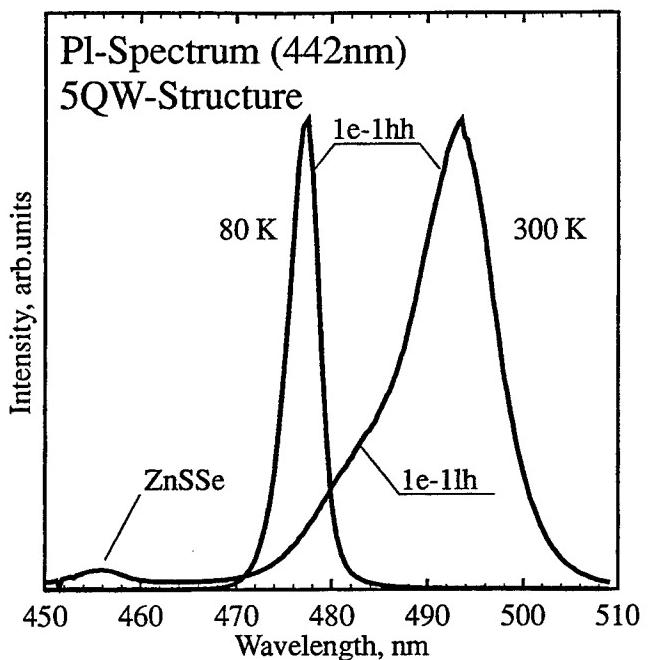


Fig. 5. Photoluminescence from Cd_{0.18}Zn_{0.82}Se quantum wells at 300 K and 80 K.

in ZnSSe. Figure 6 shows the calculated current-voltage ($I-V$) curves of the pn-junction in ZnSSe which is assumed to be grown on various III-V layers. The effects of buffer layers are remarkable. We can see that the voltage drops from about 20 V to 4 V, as the III-V layers are placed between GaAs and ZnSe. This calculation may turn out to be very useful in a later time when we attempt to prepare ZnSe-based lasers with n-on-p configuration.

4. Conclusions

We have grown ZnSe, ZnSSe, MnZnSSe and CdZnSe layers by MBE in preliminary experiments to develop materials for

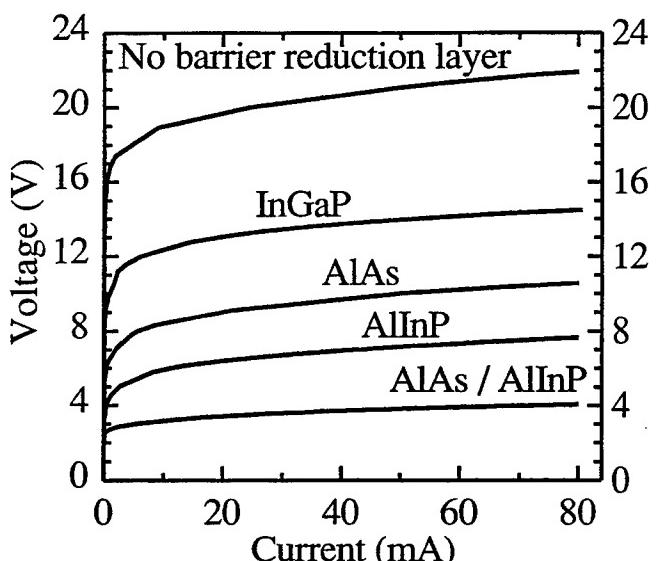


Fig. 6. Calculated current-voltage characteristics of ZnSSe pn-junctions on different barrier reduction layer combinations.

blue/green lasers and LED's. We have observed that growth on a GaAs substrate results in three-dimensional nucleation and poor structural quality, as deduced from X-ray diffraction measurements. In contrast, growth of ZnSe on a GaAs buffer layer indicates an occurrence of more two-dimensional nucleation and yields good crystal structure. Based on these observations we have used a thin ZnSe intermediate layer deposited onto a GaAs epilayer as a substrate for the subsequent growth of II-VI ternaries and quaternaries.

Of particular interest in MnZnSSe, a new quaternary the properties of which are poorly known at the moment. This material is a potential candidate for the cladding layers of blue/green lasers, perhaps capable of replacing low-quality MgZnSSe that is generally used in lasers until now.

We have proposed a use of barrier reduction layers of the III-V's on a p-GaAs substrate to alleviate the problem of injecting holes from the substrate to into the active region of a device when grown with n-on-p polarity. The calculated I-V curves for a pn-junction in ZnSSe assumed to be grown

on various buffer layers predict that a proper choice of barrier reduction layers is of very great importance.

Acknowledgement

This work was supported, in part, by the Academy of Finland. It was carried out to some extent within the framework of EU ESPRIT Smiles 8447 Program.

References

1. Park, R. M., Troffer, M. B. and Rouleau, C. M., *Appl. Phys. Lett.* **57**, 2127 (1990).
2. Haase, M. A., Qiu, J., DePuydt, J.M. and Cheng, H., *Appl. Phys. Lett.* **59**, 1272 (1991).
3. Jeon, H. *et al.*, *Appl. Phys. Lett.* **59**, 3619 (1991).
4. Gaines, J. M., Petruzzello, J. and Greenberg, B., *J. Appl. Phys.* **73**, 2835 (1993).
5. Grillo, D. C. *et al.*, *Appl. Phys. Lett.* **63**, 2723 (1993).
6. Gunshor, R. L. and Kolodziejki, L. A., *IEEE J. of QE* **24** No. 8 (1988).
7. Guha, S., Wu, B. J., Cheng, H. and DePuydt, J. M., *Appl. Phys. Lett.* **63**, 2129 (1993).

Ultrafast Nonlinear Optics in GaAs/AlGaAs Quantum Wells

J. M. Hvam, D. Birkedal, V. G. Lyssenko,¹ J. Erland² and C. B. Sørensen³

Mikroelektronik Centret, The Technical University of Denmark, DK-2800 Lyngby, Denmark

Received June 12, 1994; accepted June 15, 1994

Abstract

By degenerate four-wave mixing experiments in a two-beam geometry, we have investigated the initial coherence and dephasing of quasi two-dimensional excitons and biexcitons in GaAs multiple quantum wells. The dephasing has contributions from phonon scattering, interface-roughness scattering and exciton-exciton scattering. Inhomogeneous broadening and generation of coherent wavepackets play a significant role in the coherent exciton dynamics. The incoherent exciton dynamics, diffusion and recombination, is studied by three-beam transient grating experiments. A significant difference in the interface roughness scattering of coherent and incoherent (thermalized) excitons is observed.

1. Introduction

In low-dimensional semiconductor nanostructures as produced by modern crystal growth techniques, e.g. molecular beam epitaxy (MBE), the excitonic features of the linear and the nonlinear optical properties are strongly enhanced, and are observed even at room temperature [1, 2]. The study of exciton dynamics is of importance for applications in future ultrafast optical and opto-electronic devices, such as lasers, optical modulators and switches [3, 4]. Similarly, ultrafast linear and nonlinear optical spectroscopy of excitonic transitions provide an important tool for the study of the fundamental properties, and the quality, of these low-dimensional quantum structures. Such studies have been made possible by the development of ultrafast laser systems, and they have recently been undertaken by many research groups around the world [5–8].

We have performed coherent degenerate four-wave mixing (DFWM) experiments in a two-beam geometry on GaAs/AlGaAs multiple quantum wells, investigating the initial coherence and dephasing of quasi two-dimensional excitons after resonant excitation by ultrafast laser pulses [9]. Well width fluctuations and interface roughness lead to inhomogeneous broadening and/or splitting of the exciton lines depending on the interface island configuration [2, 10]. The formation of biexcitons in GaAs MQWs is being observed [11, 12], and in the coherent DFWM experiments a number of quantum beats and polarization interferences between different exciton transitions appear [13–17]. By spectrally resolving the DFWM signal, the degree of inhomogeneous broadening can be estimated [18] and the nature of the observed interferences and beats determined [19].

DFWM, or transient grating [20], experiments in a three-beam geometry have been performed, revealing recombination lifetimes and diffusion coefficients of the excitons.

From the diffusion coefficients, the momentum relaxation rates of the excitons are being determined and compared with dephasing, or polarization relaxation rates [21].

2. Excitons in GaAs/AlGaAs quantum wells

In bulk GaAs, the exciton binding energy is 4.2 meV with an exciton Bohr radius of 75 Å. Thus, distinct exciton features are only observable at low temperatures. The binding energy of biexcitons in bulk GaAs has been calculated to be 0.1 meV [22], so they are difficult to observe experimentally. In GaAs quantum wells with a width below the exciton diameter, there is strong confinement of the excitons in the direction perpendicular to the quantum well layers, so the excitons become quasi-two-dimensional. The forced increased overlap between the electron and hole wave functions increases the Coulomb binding energy of the excitons to about 10 meV for 100 Å quantum wells. Similarly, the biexciton binding energy is strongly enhanced and is calculated to be around 1.2 meV for 100 Å wells [23]. One therefore expects, the biexciton effects are important for the nonlinear optical properties of GaAs quantum wells, at least at low temperatures, and they have been searched for through a number of years [24–27]. It is not until recently, however, that biexcitons have been unambiguously identified by four-wave mixing and quantum beat experiments [11, 12]. We shall return to these biexciton effects later.

In bulk GaAs, the heavy hole and the light hole valence bands are degenerate at the zone centre. Since the confinement energy of free carriers in quantum wells is dependent on their mass, the degeneracy of the heavy hole and the light hole valence bands is lifted, and therefore the linear optical absorption near the band gap of GaAs quantum wells shows two distinct lines that are the heavy hole (HH) exciton and the light hole (LH) exciton resonances as shown in Fig. 1. The exciton dynamics determine the dynamical as well as the nonlinear optical properties of GaAs quantum wells, as we shall see. The optically excited excitons are created in definite k -states due to wavevector conservation in the optical absorption process. Excitation by a short coherent laser pulse, thus creates a macroscopic polarization in the medium that will decay with the scattering of excitons out of the initial k -states. This process is the dephasing of the excitons with the characteristic dephasing time, or transverse relaxation time T_2 , which is of the order of ten picoseconds at low temperature and low density. After the dephasing, energy relaxation or thermalization of the excitons will occur before the final (radiative) recombination of the excitons in about a nanosecond, depending on temperature.

With increasing excitation density, strong nonlinearities are observed in the excitonic features. Exciton collisions will

¹ Institute of Microelectronics, Chernogolovka, Moscow 142432, Russia.

² Fysisk Institut, Odense Universitet, DK-5230 Odense M, Denmark.

³ The Niels Bohr Institute, Ørsted Laboratory, Universitetsparken 5, DK-2100 København Ø, Denmark.

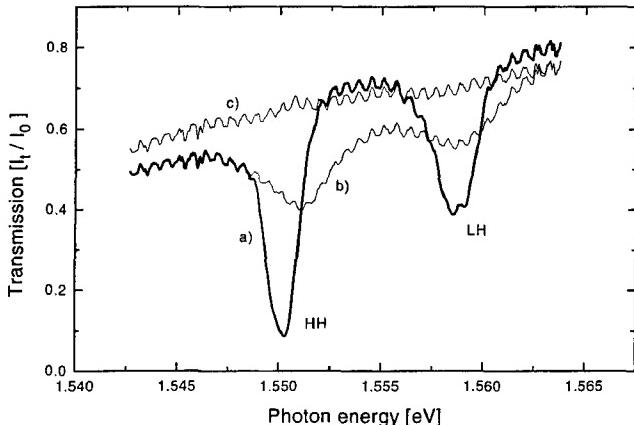


Fig. 1. Transmission through a sample with 100 Å GaAs MQWs in the region of heavy hole (HH) excitons and light hole (LH) excitons for a weak probe pulse alone (a), and after an intense pump pulse, which partially (b), or fully (c), bleaches the exciton transitions. The high frequency modulations of the spectra is due to Fabry-Perot interferences in the sample mount (sapphire).

increase the dephasing rate, resulting in broadening of the exciton lines. Furthermore, the exciton absorption is gradually saturated due to one of two causes: screening of the Coulomb interaction by free electrons (and holes) and/or phase space filling revealing the Fermionic origin of the excitons [1]. In bulk semiconductors, screening tends to be most important, but in lower-dimensional structures screening is partially inhibited and at the same time the density of states is altered, so that phase space filling becomes the dominant nonlinear effect. At higher excitation densities, the excitons are completely bleached and the band gap is renormalized due to many-body exchange and correlation effects. This is illustrated also in Fig. 1 showing absorption curves after intense pulsed excitations.

In this paper, we shall study the excitonic nonlinearities at low and medium exciton densities by degenerate four-wave mixing (DFWM), a technique that is also ideal to investigate the coherent as well as the incoherent exciton dynamics in semiconductor quantum wells. In particular, we shall study the influence of interface roughness on the exciton scattering rate and exciton motion. Interface roughness on a length scale shorter than the exciton diameter will give rise to increased scattering of electrons and holes individually, i.e. to increased dephasing, or line broadening, of the excitons. Interface roughness on a length scale longer than the exciton diameter, will affect lateral diffusion of the excitons, limit their coherence volume (area), and eventually lead to localization of the excitons. In the next section, we shall briefly describe the four-wave mixing techniques employed in this work.

3. Four-wave mixing

In general, four-wave mixing is the mixing of three input waves in a nonlinear medium, setting up a third order nonlinear polarization which in turn serves as an antenna for a fourth outgoing signal wave. In standard nonlinear optics terms, the third order nonlinear polarization is expressed as

$$\mathbf{P}^{(3)}(\mathbf{k}, \omega) = \chi^{(3)}(\omega; \omega_1, \omega_2, \omega_3) \cdot \mathbf{E}_1(\mathbf{k}_1, \omega_1) \mathbf{E}_2(\mathbf{k}_2, \omega_2) \mathbf{E}_3(\mathbf{k}_3, \omega_3) \quad (1)$$

where $\chi^{(3)}$ is the third order nonlinear susceptibility, and the frequency and the wavevector of the collimated signal beam are given by the energy and wavevector conservations, expressing the phase-matching conditions

$$\omega = \pm \omega_1 \pm \omega_2 \pm \omega_3 \wedge \mathbf{k} = \pm \mathbf{k}_1 \pm \mathbf{k}_2 \pm \mathbf{k}_3 \quad (2)$$

where $+/ -$ is entered for absorbed/emitted waves.

With two degenerate beams ($\omega_1 = \omega_2 = \omega$ and $\mathbf{k}_1 = \mathbf{k}_2 = \mathbf{k}$), impinging under an angle θ ($\cos \theta = \mathbf{k}_1 \cdot \mathbf{k}_2 / k^2 < 1$), as in Fig. 2, there are two possibilities with perfect phase-match: $\mathbf{k} = \mathbf{k}_1$ and $\mathbf{k} = \mathbf{k}_2$. These nonlinear signals are generated in a direction with a large linear signal and may therefore be difficult to detect at moderate intensities. There are also two possibilities with near phase-match: $\mathbf{k} = 2\mathbf{k}_2 - \mathbf{k}_1 - \Delta\mathbf{k}$ and $\mathbf{k}' = 2\mathbf{k}_1 - \mathbf{k}_2 - \Delta\mathbf{k}$, where $\Delta\mathbf{k}$ is the wave vector mismatch perpendicular to the sample plane. This geometry (see Fig. 2) has the advantage that the nonlinear signal is a collimated beam propagating in a background free direction. It is therefore well suited for the detection of even very small nonlinear signals. The linear background can to a high degree be eliminated by simple spatial filtering.

This DFWM can also be viewed upon as a case of light induced gratings [20, 28]. The two incident beams set up a stationary polarization grating in the nonlinear medium with scattering vectors $\pm(\mathbf{k}_2 - \mathbf{k}_1)$ and a grating constant $\Lambda = 2\pi/|\mathbf{k}_2 - \mathbf{k}_1|$. This grating in turn scatters the incident beams \mathbf{k}_1 and \mathbf{k}_2 into the directions \mathbf{k}' and \mathbf{k} , respectively, as also discussed above (see Fig. 2). The signal intensity in the scattered direction is [28]

$$I_s \propto d^2 \frac{\sin^2 \left(\frac{\Delta kd}{2} \right)}{\left(\frac{\Delta kd}{2} \right)^2} \quad (3)$$

where d is the sample thickness, or the nonlinear interaction length in general. Near-phase-match requires $\Delta kd \ll 1$, i.e. thin sample geometry. If the sample is very thin ($kd \ll 1$), the grating is essentially two-dimensional, and back scattering (with $\Delta k_b \approx 2k$) will occur with about the same intensity as in the forward direction (see Fig. 2). Note, however, that in both directions the signal will be weak, because $I_s \propto d^2$. Therefore, strong resonance enhancement will normally be necessary in order to observe a back scattered signal. On the other hand, it does open up the possibility to observe the nonlinear interaction in a resonance where strong linear

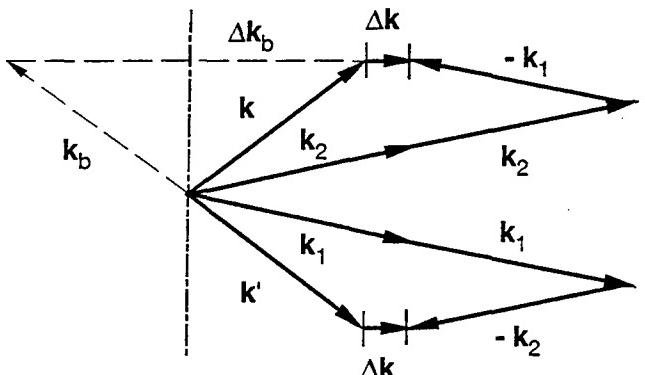


Fig. 2. Wave vector conservation for DFWM with two light beams, \mathbf{k}_1 and \mathbf{k}_2 .

absorption prevents transmission even through a thin sample [29].

With three incident light beams, more freedom is at hand to perform different types of four-wave mixing experiments. In the degenerate case $\omega_1 = \omega_2 = \omega_3$, and with three different beam directions, \mathbf{k}_1 , \mathbf{k}_2 and \mathbf{k}_3 , a certain phase mismatch $\Delta\mathbf{k}$ again requires thin samples $d \ll 1/\Delta k$, and gives rise to first order scattering ($\chi^{(3)}$) in the three directions $\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3 - \Delta\mathbf{k}$, $\mathbf{k}' = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \Delta\mathbf{k}$ and $\mathbf{k}'' = \mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 - \Delta\mathbf{k}$.

It is of course only the coherent contribution to the nonlinear signal that is well described by a nonlinear susceptibility as in eq. (1). When using DFWM experiments to measure the magnitude (in resonance) of the latter, it is therefore essential to identify the coherent contribution as for example in a transient experiment with a time resolution better than the dephasing time of the resonance. In a resonant c.w. experiment, the incoherent contribution from the photoexcited carriers integrates up over the grating lifetime, and may thus exceed by several orders of magnitude the coherent contribution.

4. Transient degenerate four-wave mixing

Performing DFWM with ultrashort laser pulses, one can obtain not only spectral information about the nonlinear coefficients, but also dynamical information about the optical excitations. The type of information obtained depends on the actual experimental configuration, as well as the character of the samples investigated. Transient DFWM is a modification of the excite-and-probe technique, where the pump consists of two beams, split off the same coherent laser beam, setting up a coherent polarization grating in the medium, and the probe is either one of the pump beams self-diffracting in this polarization (or density) grating [two-beam configuration, Fig. 3(a)] or a third delayed beam diffracting in the grating set up by the two first pulses [three-beam configuration, Fig. 3(b)]. We shall examine the two cases a little closer in the following.

4.1. Two-beam DFWM

The two incident laser pulses are split off the same laser pulse (pulse length τ_L), and are impinging on the sample with a variable optical delay between them. In order for the two laser pulses to interact coherently in the nonlinear medium, for example by setting up a polarization grating, the delay between them should not exceed the dephasing time of the nonlinear polarization in the medium, caused by the first laser pulse. The nonlinear DFWM signal is then self-diffraction of the second pulse in the grating set up by the coherent overlap between the polarizations from the first and the second pulse. For pulse 1 arriving first ($\tau_{12} > 0$) as

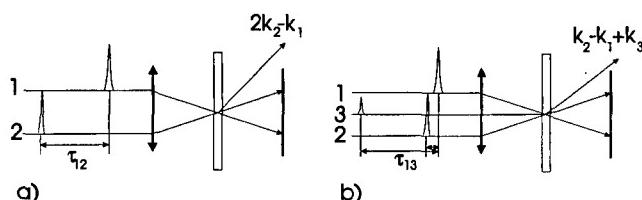


Fig. 3. Transient DFWM with two beams (a) and three beams (b), respectively.

in Fig. 3(a), a signal will thus be emitted in the direction $2\mathbf{k}_2 - \mathbf{k}_1$ as indicated. This result is obtained by solving the two-level optical Bloch equations to third order, which for δ -pulses yields [30, 31]

$$\mathbf{P}^{(3)}(\mathbf{r}, t, \tau_{12} > 0) \propto \Theta(t) \mu_{eg}^4 e^{i[(2\mathbf{k}_2 - \mathbf{k}_1)\mathbf{r} - \Omega_{eg}t + \Omega_{eg}\tau_{12}]} \quad (4)$$

where t is the time after the arrival of the second pulse, $\Theta(t)$ is the Heaviside step function, and $\Omega_{eg} = \omega_{eg} - i\gamma_{eg}$ is the complex transition frequency with the damping $\gamma_{eg} = 1/T_2 + 1/(2T_1)$. T_1 and T_2 are the longitudinal and transverse relaxation times, respectively. The intensity of the DFWM signal in the direction $2\mathbf{k}_2 - \mathbf{k}_1$ is then

$$I_{DFWM} \propto |\mathbf{P}^{(3)}(t, \tau_{12} > 0)|^2 \propto \Theta(t) \mu_{eg}^8 e^{-2\gamma_{eg}(t+\tau_{12})} \quad (5)$$

The corresponding signal in the frequency domain, as experimentally obtained by a spectrometer with a slow detector, and mathematically by Fourier transforming eq. (4) is

$$I_{DFWM} \propto |\mathbf{P}^{(3)}(\omega, \tau_{12} > 0)|^2 \propto \frac{\mu_{eg}^8}{|\Omega_{eg} - \omega|^2} e^{-2\gamma_{eg}\tau_{12}} \quad (6)$$

where ω is the optical frequency of the detected signal. These correlation traces of the time-integrated DFWM signal contain information about the dephasing rate γ_{eg} as well as the resonance enhancements of the nonlinear signal.

If the above two-level system is inhomogeneously broadened, similar expressions can be obtained by integrating over the inhomogeneous distribution of transition frequencies, e.g. $g(\omega_{eg}) \propto \exp\{-2 \ln 2 (\omega_{eg} - \omega_0)^2 / \Gamma_{eg}^2\}$, where ω_0 is the centre frequency and Γ_{eg} is the inhomogeneous intensity linewidth (FWHM).

$$\mathbf{P}_{inh}^{(3)}(t, \tau_{12}) \propto \int_0^\infty \mathbf{P}^{(3)}(t, \tau_{12}) g(\omega_{eg}) d\omega_{eg} \quad (7)$$

In this case, destructive and constructive interferences between the different transition frequencies play a significant role, and the signal in the $2\mathbf{k}_2 - \mathbf{k}_1$ direction appears as a photon echo (PE) at the time $t = \tau_{12}$.

$$I_{PE} \propto |\mathbf{P}_{inh}^{(3)}(t, \tau_{12} > 0)|^2 \propto \mu_{eg}^8 e^{-[4\Gamma_{eg}^2(t-\tau_{12})^2]/(16 \ln 2)} e^{-2\gamma_{eg}(t+\tau_{12})} \quad (8)$$

and the spectrally resolved time-integrated PE takes, for large delays ($\tau_{12} \geq 4/\Gamma_{eg}$), the form

$$I_{PE} \propto |\mathbf{P}_{inh}^{(3)}(\omega, \tau_{12} > 0)|^2 \propto \mu_{eg}^8 e^{-8 \ln 2 ((\omega_0 - \omega)^2) / \Gamma_{eg}^2} e^{-4\gamma_{eg}\tau_{12}} \quad (9)$$

which for the same homogeneous damping decays twice as fast as the free polarization decay in eq. (6). From a comparison of the observed DFWM decay with the linewidth, as observed in absorption or emission, it can be decided whether the transition is homogeneously or inhomogeneously broadened. Hence, the dephasing time, and thereby the homogeneous linewidth, can be determined from either eq. (6) or eq. (9).

4.2. Three-beam DFWM

This situation is sketched in Fig. 3(b). The first two pulses arrive simultaneously, or well within the dephasing time of the material ($\tau_{12} \ll T_2$), and interfere coherently to set up a nonlinear grating in the medium. This grating can then be detected at variable time delays, τ_{13} , by diffraction of the third pulse. If also $\tau_{13} \ll T_2$, then pulse 3 will diffract off a

coherent polarization grating set up by pulse 1 and pulse 2, as in the self-diffraction case above. If, however, ω is in resonance with an electronic excitation in the material, a real excitation density grating may persist in the material long after the coherent polarization grating has disappeared by dephasing. This grating, however, still bears the fingerprint of the coherent overlap between the two first pulses, and can even give rise to a photon echo signal, stimulated by the third pulse, but dependent on the delay τ_{12} . If, on the other hand, the scattered signal is recorded as a function of delay τ_{13} of the third pulse, the decay of the incoherent excitation population is being monitored. This type of experiment is therefore well suited to separate the coherent contribution to the optical nonlinearities from the more long-lived incoherent contributions from a high density of excited carriers in the medium.

The lifetime τ_G of the incoherent excitation density grating is determined by the lifetime τ_R of the excited carriers as well as by carrier diffusion [32] since the latter will wash out the spatial modulation of the carrier density.

$$\tau_G^{-1} = \tau_R^{-1} + \frac{4\pi^2 D}{\Lambda} \quad (10)$$

where D is the carrier diffusion coefficient and $\Lambda = 2\pi/|\mathbf{k}_2 - \mathbf{k}_1| = \lambda/[2 \sin(\theta/2)]$ is the grating constant, as determined by the wavelength λ of the exciting light and the angle θ between the two interfering beams (\mathbf{k}_1 and \mathbf{k}_2). The grating lifetime is determined from the decay of the integrated intensity of the scattered test signal as a function of the delay τ_{13} of the test pulse (3). Hence, the carrier lifetime and the diffusion coefficient can be determined separately from eq. (33) by performing transient experiments at different angles θ .

4.3. Experimental setup

A typical experimental setup to perform transient DFWM is shown in Fig. 4. The output from a synchronously pumped dye-laser system, or a self mode-locked Ti : sapphire laser, is divided into two or three beams with computer controlled delays in two of the beams. The samples are placed in a variable temperature liquid-helium cryostat and the output signal beam is selected, dispersed in a spectrometer, and measured time integrated by an optical multichannel analyser system. The simultaneous spectral resolution of the output signal and temporal resolution of the input signals

yield, together with the energy conservation and the resonance enhancements of the nonlinear mixing process, the maximum information, allowed by the uncertainty principle, about the resonances involved in the FWM process.

5. Coherent exciton dynamics

Figure 5 shows an example of the coherent exciton dynamics as revealed by a two-beam transient DFWM experiment in a specially prepared MBE-grown sample with a series of multiple quantum well (MQW) structures (10 periods) with well widths $L_z = 80 \text{ \AA}$, 100 \AA , 130 \AA , and 160 \AA and $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barriers of width 150 \AA . From the DFWM spectra recorded as a function of the delay between the input pulses, a number of observations are made:

(1) The strong resonance enhancement of the nonlinear signal (see eqs (6) and (9)) around the fundamental excitonic resonances, heavy hole excitons (HH) and light hole excitons (LH), is clearly observed for the different MQW series.

(2) The asymmetric behaviour of the DFWM signal as a function of the delay is in agreement with the simple two-level models in eqs (6) and (9) and should reveal the dephasing time $T_2 = 1/\gamma_{eg}$ of the corresponding transitions (see below) [30,31].

(3) A signal for negative delays is a signature of polarization, or exciton, interactions [33, 34] and is clearly seen for the narrower wells. In particular for the 100 \AA wells, is seen a line (XX) that is dominant for negative delays. This line is due to two-photon transitions to the biexciton state associated with the heavy hole exciton resonance HH_{100} of the 100 \AA quantum wells [7], as can be ascertained by polarization resolved experiments. The shift of the XX line with respect to the HH line is equal to the biexciton binding energy [12].

(4) The pronounced modulation observed on the HH_{100} line is due to a quantum beat between the HH_{100} and the LH_{100} resonances [13–15]. The HH excitons and the LH excitons within the same MQW series are created in a coherent superposition state by the spectrally broad laser pulse, and the resulting polarization oscillates in time with a frequency $\Delta\omega = \Delta E/\hbar$, where ΔE is the splitting of the HH and LH excitons. As a result, the DFWM signal as a function of delay oscillates with the period $2\pi/\Delta\omega$ [31]. Beats between HH excitons and biexcitons have also been observed, confirming the formation of biexcitons in the GaAs quantum wells [11, 12].

(5) In Fig. 5, the laser line is centred around the HH_{100} line, but is sufficiently broad to excite all the resonances

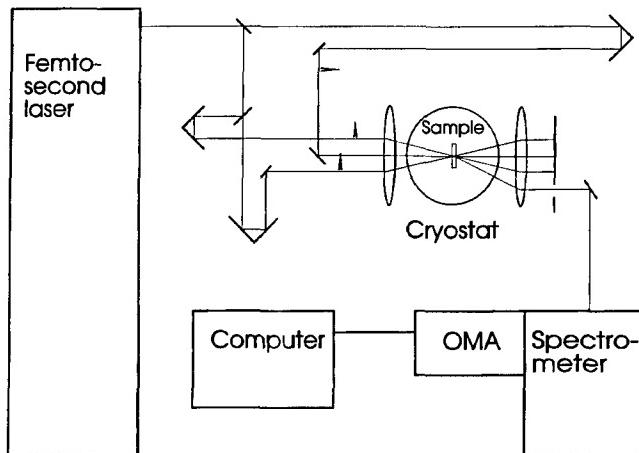


Fig. 4. Experimental setup.

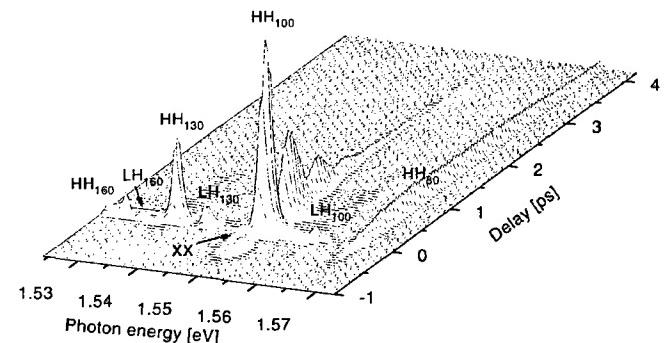


Fig. 5. Transient DFWM spectra as a function of delay between the input pulses for GaAs MQWs of different widths, see discussion in text.

observed. The DFWM signal from the wider well resonances decay much faster than the signal from the narrower wells, suggesting shorter dephasing times for the wider wells. One reason could be that the wider wells are excited well above the resonances, so that higher excited exciton states and continuum states contribute to the dephasing of the HH and LH excitons. Such a conclusion, however, could be wrong, because a corresponding broadening of the resonances is not observed in Fig. 5. Instead we conclude that the fast decay of the signal is due to destructive interference between the $n = 1$ HH (LH) excitons and the higher excited exciton states and continuum states [35]. A wave packet is formed as a coherent superposition of many excited states, and the corresponding polarization dies out in a time proportional to $1/(\text{spread in energy of excited states})$. Note that no rephasing is initiated by the second pulse, as in the case of a coherently excited inhomogeneous distribution of transition frequencies (photon echo, see the discussion below) [18, 30, 31].

(6) In Fig. 5, the HH_{100} excitons are the only ones excited alone without a significant contribution from LH excitons and higher excited states. Therefore the observed decay is representative of the dephasing of HH excitons in GaAs MQWs at low density and low temperature. The observed dephasing time is about 10 ps, and the fact that the signal is maximum for a finite positive delay indicates that in this case the transition is inhomogeneously broadened and the signal appears as a photon echo [18].

The observed dephasing times T_2 are slightly larger than in bulk GaAs [16], but the temperature dependence is relatively weak. However, it turns out that there is a significant difference in the scattering rates, that are measured in a coherent DFWM experiment [9], and those that are deduced from measuring exciton diffusion in the two-dimensional GaAs layers [21], as will be discussed in the next section.

6. Incoherent exciton dynamics

As mentioned in Section 4.2, the diffusion coefficient D and the recombination lifetime τ_R of the two-dimensional excitons can be determined from a three-beam transient grating experiment. Figure 6 shows the measured temperature dependence of D and τ_R [21]. The solid curve is a fit, to be

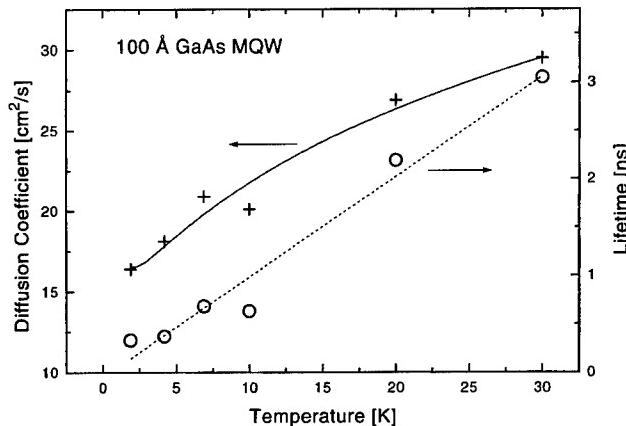


Fig. 6. Diffusion coefficient of excitons in GaAs 100 Å quantum wells (+) vs. sample temperature. The solid curve represents a fit to the scattering rates (eq. (12)) in Fig. 7. Also shown is the temperature dependence of the exciton lifetime (○) with a linear fit (dashed line).

discussed below. The dashed line is a linear fit to the exciton lifetime as a function of temperature with a slope of 100 ps/K. The lifetime increases due to increasing thermal population of nonradiative states. Only excitons with an in-plane wavevector smaller than the wave vector of a photon with the transition energy may decay radiatively.

The momentum relaxation rate $1/\tau_p$ can be deduced from the diffusion constant applying the classical expression

$$D = \frac{kT}{m_x} \tau_p \quad (11)$$

where m_x and τ_p are the effective mass and the momentum relaxation time for the exciton, respectively. Figure 7 shows the temperature dependence of the momentum relaxation rate $1/\tau_p$ (circles) and the dephasing rates $1/T_2$ (squares). At low temperature, both rates are about 50 ns^{-1} . The dashed line is a fit obtained from linear regression to the dephasing results. The slope is $(2.6 \pm 0.7) \times 10^9 \text{ s}^{-1} \text{ K}^{-1}$. The low temperature limit is 40 ns^{-1} representing the residual interface-roughness scattering. The momentum relaxation rate increases much faster with temperature and shows a nonlinear dependence. The solid curve is a fit by the formula

$$\frac{1}{\tau} = C + \frac{1}{\tau_{ir}} + \frac{1}{\tau_{ph}} = C + \frac{R\sqrt{T}}{L_z^2} + \frac{AT}{L_z} \quad (12)$$

where τ_{ir} and τ_{ph} are the times for interface-roughness and phonon scattering, respectively, T is the temperature and L_z is the well width. The fitting parameters are $A = 54 \text{ ms}^{-1} \text{ K}^{-1}$, $R = 9 \times 10^{-6} \text{ m}^2 (\text{s}\sqrt{T})^{-1}$, and $C = -6.6 \times 10^{10} \text{ s}^{-1}$.

The expression for the phonon scattering is intuitively easily understood. It is proportional to the temperature due to the classical phonon population, and inversely proportional to the well width, as is the density of exciton states to scatter into. The expression for the interface-roughness scattering is also explained classically [21, 36]. The free excitons are considered to scatter off islands of slightly different well width (about one monolayer). The mean free path of excitons is assumed to be independent of temperature. Hence, the rate of scattering will be proportional to the exciton velocity (\sqrt{T}) and the height of the scattering potential ($1/L_z^2$). From the experimental results, it seems that this type of interface-roughness scattering is not effective for

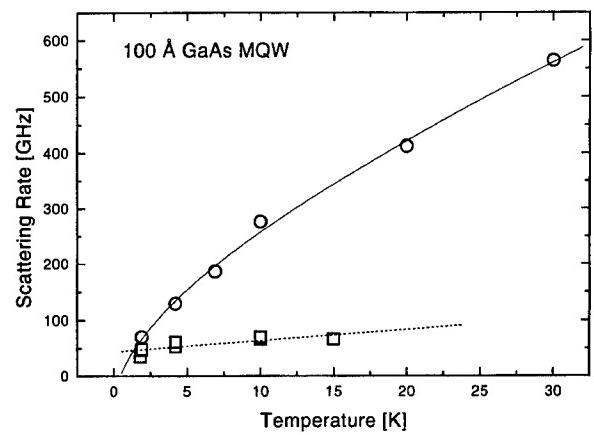


Fig. 7. Scattering rate of the thermalized excitons (○) with second order polynomial fit in \sqrt{T} (solid curve) according to eq. (12). Scattering rates of coherent excitons from dephasing measurements (□) with a linear fit (dashed curve).

dephasing. This, however, is not too surprising, because the coherent excitons are formed with negligible wave vector along the two-dimensional layers. The thermalized excitons involved in the diffusion, on the other hand have a significant in-plane component of their wave vector.

7. Conclusions

We have found that transient DFWM is a very powerful tool to investigate the ultrafast linear and nonlinear optical processes in semiconductor quantum structures. Fundamental parameters for the coherent and incoherent exciton dynamics can be extracted from such measurements. Care has to be taken, however, in the interpretation of the observed results. Quantum interferences and polarization interferences may play a significant role in the observed dynamics of the macroscopic polarization. In particular, inhomogeneous broadening and the coherent excitation of continuum wave packets has to be taken into account.

It is concluded that GaAs MQWs show strong nonlinearities in the excitonic transitions and that the coherent response times of these transitions is in the picosecond or subpicosecond range. Nonlinear optical devices based on excitonic transitions in low-dimensional semiconductor nanostructures are therefore promising as ultrafast electro-optical or all-optical devices.

Acknowledgements

Thanks are due to I. Balslev for many helpful discussions. The work was supported by the Danish Natural Science Research Council.

References

1. Chemla, D. S. and Miller, D. A. B., *J. Opt. Soc. Am.* **B2**, 1155 (1985).
2. Göbel, E. O. and Ploog, K., *Prog. Quant. Electr.* **14**, 289 (1990).
3. Knox, W. H., *Appl. Phys.* **A53**, 503 (1991).
4. Hvam, J. M., Erland, J. and Pantke, K-H., *SPIE Proc. Vol. 2142* (1994) in press.
5. Weiner, A. M., Silvestri, S. and Ippen, E. P., *J. Opt. Soc. Am.* **B2**, 654 (1985).
6. Schultheis, L., Kuhl, J., Honold, A. and Tu, C. W., *Phys. Rev. Lett.* **57**, 1635 (1986).
7. Dörnfeld, C. and Hvam, J. M., *IEEE J. Quantum Electron.* **QE-25**, 904 (1989).
8. Göbel, E. O., Leo, K., Damen, T. C. and Shah, J., *Phys. Rev. Lett.* **64**, 1801 (1990).
9. Oberhauser, D. et al., *Proc. Int. Conf. Laser '91*, San Diego (1991), p. 368.
10. Kopf, R. F., Schubert, E. F., Harris, T. D. and Becker, R. S., *Appl. Phys. Lett.* **58**, 631 (1991).
11. Lovering, D. J., Phillips, R. T., Denton, G. J. and Smith, G. W., *Phys. Rev. Lett.* **68**, 1880 (1992).
12. Pantke, K-H., Oberhauser, D., Lyssenko, V. G., Hvam, J. M. and Weimann, G., *Phys. Rev.* **B47**, 2413 (1993).
13. Leo, K., Damen, T. C., Shah, J., Göbel, E. O. and Köhler, K., *Appl. Phys. Lett.* **57**, 19 (1990).
14. Feuerbacher, B. F., Kuhl, J., Eccleston, R. and Ploog, K., *Solid State Commun.* **74**, 1279 (1990).
15. Leo, K. et al., *Phys. Rev.* **B44**, 5726 (1991).
16. Leo, K., Damen, T. C., Shah, J. and Köhler, K., *Phys. Rev.* **B42**, 11359 (1990).
17. Koch, M. et al., *Phys. Rev.* **B48**, 11480 (1993).
18. Erland, J., Pantke, K-H., Mizeikis, V., Lyssenko, V. G. and Hvam, J. M., to be published.
19. Lyssenko, V. G. et al., *Phys. Rev.* **B48**, 5720 (1993).
20. Eichler, H. J., Günter, P. and Pohl, D. W., "Laser Induced Dynamical Gratings", Springer Ser. Opt. Sci. **50** (Springer, Berlin 1986).
21. Oberhauser, D., Pantke, K-H., Hvam, J. M., Weimann, G. and Klingshirn, C., *Phys. Rev.* **B47**, 6827 (1993).
22. Brinkman, W. F., Rice, T. M. and Bell, B., *Phys. Rev.* **B8**, 1570 (1973).
23. Kleinman, D. A., *Phys. Rev.* **B28**, 871 (1983).
24. Miller, R. C., Kleinman, D. A., Gossard, A. C. and Munteanu, O., *Phys. Rev.* **B25**, 6545 (1982).
25. Cingolani, R., Chen, Y. and Ploog, K., *Phys. Rev.* **B38**, 13478 (1988).
26. Charbonneau, S. et al., *Phys. Rev.* **B38**, 3583 (1988).
27. Phillips, R. T., Lovering, D. J., Denton, G. J. and Smith, G. W., *Phys. Rev.* **B45**, 4308 (1992).
28. Maruani, A. and Chemla, D. S., *Phys. Rev.* **B23**, 841 (1981).
29. Hvam J. M. and Dörnfeld, C., in "Optical Switching in Low-Dimensional Systems" (Edited by H. Haug and L. Banyai) (Plenum Publishing Corp., New York 1989), p. 233.
30. Yajima, T. and Taira, Y., *J. Phys. Soc. Japan* **47**, 1620 (1979).
31. Erland, J. and Balslev, I., *Phys. Rev.* **B48**, 1765 (1994).
32. Moss, S. C., Lindle, J. R., Mackey, H. J. and Smirl, A., *Appl. Phys. Lett.* **39**, 227 (1981).
33. Leo, K. et al., *Phys. Rev. Lett.* **65**, 1340 (1990).
34. Feuerbacher, B. F., Kuhl, J. and Ploog, K., *Phys. Rev.* **B43**, 2439 (1991).
35. Feldman, J. et al., *Phys. Rev. Lett.* **70**, 3027 (1993).
36. Tsien, K. T., Sankey, O. F. and Morkoç, H., *Appl. Phys. Lett.* **57**, 1666 (1990).

Modulation- and Transmission-Ellipsometric Characterization of Semiconductor Heterostructures

Krikor B. Ozanyan, Turid Worren and Ola Hunderi

Department of Physics, Norwegian Institute of Technology, N-7034 Trondheim, Norway

Received May 26, 1994; accepted in revised form July 5, 1994

Abstract

The potential of spectroscopic ellipsometry for characterization of III-V semiconductor quantum heterostructures is demonstrated by two ellipsometric techniques, chosen to match polarization selection rules and spectral region of interest. Brewster-angle geometry and transmission arrangement has been used for non-destructive studies of electronic intersubband transitions in the MIR. The inversion problem is solved to determine the extraordinary dielectric response in the wells. Photo-modulated spectroscopic ellipsometry utilizing a tunable laser source has been used for studies of near band-gap transitions in the NIR/VIS. This provides information that is not supplied by conventional techniques like PLE.

1. Introduction

Quasi-2D confined systems manifest well-defined polarization selection rules for optical transitions [1], a fact now confirmed by a variety of techniques. Polarization sensitivity is increasingly introduced in experiments, that traditionally have not been aimed at acquiring such information. Polarization-sensitive photoluminescence excitation is one example, successful in distinguishing between electron-heavy-hole (e-hh) and electron-light-hole (e-lh) transitions in quantized structures.

Ellipsometry offers a natural way to study optical transitions with pronounced polarization dependence by measuring the ellipsometric angles Ψ and Δ for the experimental configuration of interest. The squared modulus $\tan^2(\Psi)$ of the polarization ratio

$$\chi = \tan \Psi e^{i\Delta} = \frac{\chi^O}{\chi^I} = \frac{E_p^O/E_s^O}{E_p^I/E_s^I}, \quad (1)$$

(subscripts stand for the polarization directions of the incident I and outgoing O electric field amplitudes) yields the detected p-polarized intensity normalized to the s-polarized one. Therefore, in addition to the information extracted from conventional transmission/reflection experiments, because of the independently measured phase angle Δ , spectroscopic ellipsometry allows the conversion to spectra of the real and imaginary parts of the dielectric response function without resorting to Kramers-Kronig analysis.

The present work deals with the potential of spectroscopic ellipsometry for characterization of optical transitions in heterostructures containing layers, thin enough to manifest quantization effects because of discontinuities in the band-edge potential. Transitions between the first and second upper-lying level emerging from the quantization of a single allowed-band (inter-subband transitions) usually occur at energies much lower than the bandgaps of the materials building the heterostructure and for the most III-V materials lie in the MIR/FIR spectral region. Tran-

sitions between electron-band levels and hole-band levels (interband transitions) involve energies from the order of the bandgap and are observable for the same materials in the NIR/Visible region. Different parts of the optical spectrum require different techniques and instrumentation, making a unique approach impossible. As revealed further, we use FTIR ellipsometry in the $10\text{ }\mu\text{m}$ region and modulation ellipsometry for wavelengths around $1\text{ }\mu\text{m}$, for monitoring intersubband and interband transitions, respectively, utilizing the appropriate radiation sources, polarizers and detectors.

We calculate the energies and the probability densities for the wavefunctions of the quantized electron and hole levels by finding [2] exact solutions of the Schrödinger equation by means of a quantum-mechanical transfer matrix method [3, 4].

2. Intersubband transitions

Intersubband transitions in heterostructures occur between quantized levels, originating from a common 3D allowed-band. This induces large matrix elements for electric field polarized parallel to the sample surface's normal (π -polarization) [1].

To achieve a considerable π -component of the exciting electric field, a beam with a considerable p-polarized component should be used at a large angle of incidence. For a refractive index between 3 and 4, typical for most III-V semiconductors, the Brewster angle is large as well, and the conventional ellipsometric approach will suffer from low intensity of the p-polarized beam, reflected from the ambient-overlayer surface. If one moves away from the conventional reflection geometry however, this may be turned into an advantage because of the more efficient coupling of the p-polarized beam into the sample and out of it, resulting in higher intensity on the detector. In search for suitable geometries a transmission arrangement is a nondestructive option, and the backside-reflection arrangement becomes favourable if a metal reflector is evaporated on the overlayer. Both arrangements require a transparent substrate, which is fulfilled for undoped GaAs in the $10\text{ }\mu\text{m}$ spectral region. The modelling of the ellipsometric signal in both cases is facilitated by choosing the incidence angle equal to the Brewster angle. We introduce the model [5] of partial coverage of the substrate, in which the existence of an overlayer is not taken into account for multiple reflections. This is justified by the fact that the p-polarized beam, probing the intersubband absorption, is coupled directly out and is not multiply reflected in the substrate. The model relates the

polarization ratio χ of the ambient-overlayer-substrate system only to the polarization ratio χ' measured by the ellipsometer:

$$\chi' = C\chi \quad (2)$$

where the factor C accounts for the substrate-ambient interface and phase-averaging in the optically thick substrate. It depends only on the substrate permittivity ϵ and is typically a constant over the range where intersubband absorption occurs.

The effect of the uniaxial overlayer is contained in χ and is modelled by transfer matrix calculations [3]. This latter approach is very similar to that mentioned in Section 1, because of the widely used analogy between optics and wave mechanics. In the case when the total thickness of the overlayer D is much smaller than the wavelength λ , linear approximation for the matrix product can be used and Ψ and Δ spectra can be analytically inverted to those of the real and imaginary part of the overlayer extraordinary dielectric function, if the ordinary one is considered known. The spatially averaged ordinary ϵ^o and extraordinary ϵ^e permittivities are given by the effective medium theory [6]

$$\epsilon^o = \langle \epsilon^o \rangle = f\epsilon_B + (1-f)\epsilon_{2D} \quad (3)$$

$$\frac{1}{\epsilon^e} = \left\langle \frac{1}{\epsilon^e} \right\rangle = \frac{f}{\epsilon_B} + \frac{1-f}{\epsilon_{QW}} \quad (4)$$

where the barrier material with filling factor f is considered isotropic ($\epsilon_B^o = \epsilon_B^e = \epsilon_B$) and the ordinary QW permittivity ϵ_{2D} is that of a 2D-plasma. Using (4) an analytical expression can be obtained for the extraordinary dielectric response of the QW layer, $\epsilon_{QW} = \epsilon_{QW}(\chi', C, D, \epsilon^o, \epsilon, \lambda)$. If linearization is not justified, fitting of the experimental Ψ and Δ spectra with the calculated ones can be applied.

The samples used for our measurements were grown by MBE and consist typically of 50 periods of GaAs/AlGaAs with alloy composition and well/barrier thickness chosen carefully to bring the separation between the first (e1) and second (e2) quantized levels in the range around 100 meV, interesting for IR detector applications. Population of the lower level e1 with electrons was achieved by doping the well layers with Si at the level of 10^{18} cm^{-3} . Measurements were performed in the range $400\text{--}4000 \text{ cm}^{-1}$ with a FTIR ellipsometer described elsewhere [7], equipped with a globar source, wire-grid polarizers and a DGTS detector. The measurements reported here were performed at room temperature.

2.1. Brewster angle transmission ellipsometry

In a transmission ellipsometric arrangement, the square of the modulus of the transmission polarization ratio τ yields the normalized transmittance spectrum,

$$|\tau|^2 = \tan^2 \psi = \frac{T_p}{T_s}, \quad (5)$$

as it would be supplied by other techniques that sense only the amplitude of the outgoing beam.

Figure 1 shows the result of the data treatment for the case $D/\lambda \ll 1$, in which the linear approximation allows us to obtain analytically numerical values for ϵ_{QW} . The substrate permittivity ϵ was measured by conventional reflec-

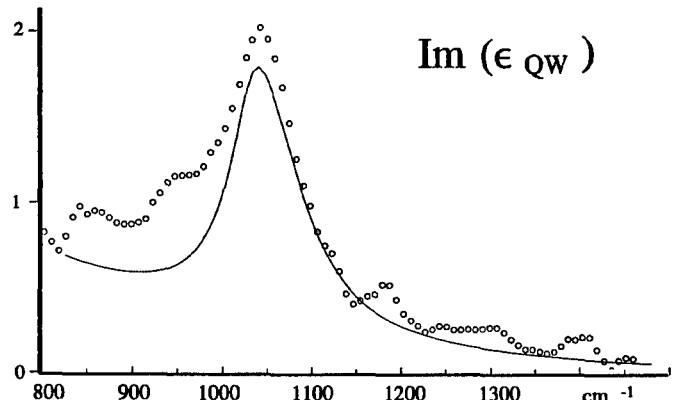


Fig. 1. Imaginary part of the extraordinary QW dielectric function, extracted from the experiment (points) and calculated within a Lorentzian oscillator model (full curve).

tion ellipsometry from the backside of the same sample. The plasma frequency was calculated assuming a free-carrier concentration inferred from the growth conditions. Barrier permittivity was calculated within a simple oscillator model, using literature data for the high-frequency dielectric constant, oscillator strength, broadening factor and oscillation frequency of the last phonon band before the region of interest. The experimentally determined spectrum of ϵ_{QW} corresponds to a Lorentzian oscillator with a center frequency 1045 cm^{-1} and broadening 100 cm^{-1} , as expected from the sample design. The particular sample reported in Fig. 1 was grown in a Varian GEN II modular MBE machine on a semi-insulating undoped GaAs (001) substrate. Growth was started with a $3 \mu\text{m}$ thick undoped GaAs buffer layer. Then followed a 60 period QW stack, each period consisting of $130 \text{ \AA Al}_{0.3}\text{Ga}_{0.7}\text{As}$ and 88 \AA GaAs layers with the middle 65 \AA of the GaAs wells n-type doped (10^{18} cm^{-3} Si). Growth was terminated with a $130 \text{ \AA Al}_{0.3}\text{Ga}_{0.7}\text{As}$ barrier and a 50 \AA GaAs capping layer.

In addition to the transmission arrangement reported here, we have also used a backside reflection ellipsometry arrangement with a metal film evaporated on the overlayer. This geometry is more sensitive to absorption of the π -component of the electric field, due to the so-called IR selection rule. It is, however, a "destructive" technique and it will not be reported further here.

3. Interband transitions

It has been demonstrated recently that close to band-edge energies, a combination between ellipsometric spectra taken with and without modulation of an applied electric field can yield geometric parameters, the static dielectric function, the energies of the quantized transitions and their polarization dependence [3]. We use this approach combining an optimized version of conventional reflection ellipsometry with the photo-modulation technique. Spectroscopic ellipsometry is quite sensitive in quality to the light source, since the latter determines the optical signal level on the detector and the accuracy with which the angle of incidence is known. The use of conventional illumination sources with a low spectral brilliance results in low intensity after monochromatization and if the probe beam is not perfectly collimated, measuring and taking into account beam divergence has to

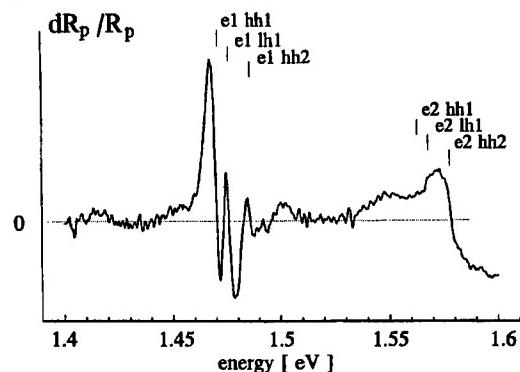


Fig. 2. dR_p/R_p spectra for a p-i-n test sample with a MQW structure as an i-region. Calculated transitions are marked on top.

be included. We believe to have solved those problems by using a Ti : Sapphire laser, tunable in the range 1.2–1.8 eV, covering the bandgap energies of the mostly used III-V binary semiconductors and solid-solutions. The visible part of this range can be used for alignment and calibration which thus become considerably simpler and better in precision. With an output power of 0.1–0.2 W at large angle of incidence the optical signal level is high enough to be attenuated within orders of magnitude to match the best detector performance (linearity, sensitivity, etc.). The intensity measured by the ellipsometer with one of the polarizers fixed at 45° and the other at angle ω with respect to the incidence plane is given by [9]:

$$I(\omega) = \gamma + \alpha \cos 2\omega + \beta \sin 2\omega, \quad (6)$$

where the coefficients γ , $-\alpha$ and β are the first three components of the Stokes vector. The usual experimental procedure is to rotate the analyzer continuously or stepwise and obtain the coefficients by using a two-phase lock-in technique or Fourier analysis, respectively. The expected improvement in the signal-to-noise ratio in our case allows a simpler and quicker modification, known as Beattie's method [10]. With the polarizer kept at 45° we measure the intensity at four values for the angle between the polarizer and analyzer positions:

$$\begin{aligned} I(0^\circ) &= \frac{I_0}{2} R_p \\ I(90^\circ) &= \frac{I_0}{2} R_s \\ I(45^\circ) &= \frac{I_0}{4} (R_s + R_p + 2\sqrt{R_s R_p} \cos \Delta) \\ I(135^\circ) &= \frac{I_0}{4} (R_s + R_p - 2\sqrt{R_s R_p} \cos \Delta) \end{aligned} \quad (7)$$

The system (7) overdetermines the unknowns R_s , R_p and Δ and the fourth measurement is used as a consistency check. When modulation is applied, the solutions will give the in-phase changes in the s and p-polarized reflectances dR_s , dR_p and in the phase difference $d\Delta$.

The experimental setup contains Glan-Thompson or Nicol-prism polarizers, mounted on precision goniometers. Measurements at different incidence angles can be performed in the temperature range 12 K–300 K by using a

closed-cycle refrigerator. The measured signal is normalized towards the spectral sensitivity of the Si-diode photodetector. The intensity of the probe beam is mechanically modulated at 480 Hz and standard lock-in technique is applied. Photo-modulation is achieved by exciting nonequilibrium free carriers close to the sample surface by illumination with above-bandgap light (He-Ne laser, 1 mW output power). The modulation frequency is 40 Hz which is a compromise between arguments for reducing experimental duration and supplying sufficient time for the space charge to reach quasiequilibrium.

The comparison with results obtained from non phase-sensitive modulation experiments, is again straightforward, since the measurements at 0° and 90° (see (7)) are equivalent to measuring the differential reflectances at s and p-polarization by photo-modulation spectroscopy [11].

Figure 2 shows the dR_p/R_p spectrum at 300 K for a test sample containing 8 periods of 100 Å GaAs wells and 100 Å $Al_{0.32}Ga_{0.68}As$ barriers at the i-region of a p-i-n structure [3]. The calculated position of some interband transitions is depicted. It can be speculated that comparison of experimental results obtained by different techniques employing the same laboratory equipment is advantageous with respect to some systematic errors. This can be the case with modulated ellipsometry and photoluminescence excitation spectra for example, since we use essentially the same equipment for both setups.

Results from modulation ellipsometry can also be compared with the FTIR ellipsometry spectra. The energy difference of the $e1-hh1$ and $e2-hh1$ structures in the photo-modulated spectrum will correspond to the energy position of the $e1-e2$ line measured with the transmission ellipsometer. Figures 1 and 2 show results from two different samples, and a direct comparison is not possible here. Modulation ellipsometry, in the spectral range reported here, will therefore also yield the separation of the $e1$ and $e2$ electronic levels, but not directly the line-shape of the $e1-e2$ intraband transition, as the two different techniques give access only to the lineshapes of the different transitions they involve.

In conclusion, we have shown some applications of spectroscopic ellipsometry for studying optical transitions in semiconductor quantum heterostructures. In addition to the information obtainable from non phase-sensitive techniques, the real and imaginary parts of the dielectric response function can be obtained analytically or numerically without Kramers-Kronig analysis.

Acknowledgements

Thanks are due to Thomas Zettler (TU Berlin) for the GaAs/AlGaAs test-sample and Trond Westgaard for participation in the reflection experiments. This work has been partially supported by the Research Council of Norway.

References

1. Weisbuch, C. and Vinter, B., in "Quantum Semiconductor Structures" (Academic Press, San Diego 1991), p. 57.
2. Ozanyan, K. B., Worren, T. and Hunderi, O. (unpublished).
3. Zettler, J.-Th., Mikkelsen, H., Leo, K., Kurz, H., Carius, R. and Förster, A., Phys. Rev. B46, 15955 (1992).

4. Lui, W. W. and Fukuma, M., *J. Appl. Phys.*, **60**, 1555 (1988).
5. Ozanyan, K. B., Hunderi, O. and Finland, B.-O., *J. Appl. Phys.*, **75**(10), 15 May 1994 (in print).
6. Hunderi, O., *J. Wave-Material Interaction* **2**, 29 (1987).
7. Bremer, J., Hunderi, O., Kong Fanping, Skauli T. and Wold, E., *Applied Optics* **31**, 471 (1992).
8. Kane, M. J., Emeny, M. T., Apsley, N., Whitehouse, C. R. and Lee, D., *Semicond. Sci. Technol.* **3**, 722 (1988).
9. Azzam, R. M. A. and Bashra, N. M., in "Ellipsometry and Polarized Light" (North-Holland Publ. Co., 1977).
10. Beattie, J. R., *Phil. Mag.* **46**, 235 (1955).
11. Pollak, F. H. and Shen, H., *J. Electron. Mater.*, **19**, 399 (1990).

All-Optical Bistability in Luminescence of Thin CdS Films

Bruno Ullrich and Takayoshi Kobayashi

Department of Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

Received May 4, 1994; accepted June 16, 1994

Abstract

Recently, we have discovered that the excitation of thin CdS films with the 514.5 nm line causes an all-optical bistable luminescent emission in the near infrared part of the spectrum. A first insight in mechanism and origin of all-optical bistability in luminescence of semiconductors is presented. We show that the initial and final states of the bistable luminescence are determined by the thermal properties of the luminescence intensity of the thin CdS film. However, during the bistable switch of the luminescence thermal effects and electronic ones take place. Thus, all-optical bistability in luminescence is a thermally induced electronic bistability.

1. Introduction

The impact of photonics on telecommunications increases rapidly since the development of the first transatlantic light-wave system in 1988. Without any doubt, the next century is expected to become an Optopia where extensive progresses and applications in photonics will be attained [1]. On the other hand, after the first euphoria surrounding all-optical bistability (all-OB) and optical computing it became quite clear that the breakthrough to a true mass-market for photonics requires much more efforts in basic research as assumed during the last decade. Furthermore, the future of photonics will depend strongly on the development of cost-effective technologies for optical components and optical interconnects, especially for shorter distance applications. Notable efforts in basic research concerning all-OBs of thin semiconducting films were reported last year by the discovery of all-OB in luminescence (all-OBL) in thin CdS films [2, 3]. In particular, we pointed out that the excitation of thin CdS films with the 514.5 nm line of an argon laser causes, besides the known photo-thermal all-OB in transmission (all-OBT) [4], highly contrasted bistable loops in the luminescence radiation. Since the radiative recombination is an electronic process whose temporal response depends on the lifetime of the excited carriers, the interesting question has arisen whether the bistable switch of all-OBL takes place faster than the switch in transmission which is a predominantly thermal effect [2, 3]. Indeed, we show in this publication that all-OBLs are *photo-thermally induced electronic bistabilities*. In particular, the contrast – i.e., the luminescence intensity of the initial and final states of the bistable loop – and sense of all-OBL are determined by the dependence of the luminescence intensity on temperature ($I_{\text{Lum}}(T)$) but during the bistable switch of all-OBL electronic processes take place.

2. Experiment

The sample investigated was a thin (0.45 μm) CdS film on an indium/tin oxide (ITO) prepared by spraying with 10^{-7} M Cu_2^+ in the solution. The spray deposition of CdS has been

described in detail in a previous work [5]. The excitation of the sample was performed with the 514.5 nm Ar^+ laser line focused to a diameter of 400 μm. The power modulation of the incident beam was performed by a liquid crystal polarizer system. The application of spray deposition results in perpendicular to the surface oriented CdS films [5]. Hence, the absorption edge is shifted to lower energies with respect to the parallel oriented material and it is necessary to cool the CdS film in order to observe all-OBs with the 514.5 nm line [4]. Therefore, the sample was mounted on a copper holder in a liquid helium contact gas cryostat. Both the transmitted and the incident power were measured with Si-photodiodes. For the luminescence measurements, a 10 cm-grating monochromator (resolution 3 nm) with a photomultiplier was placed behind the cryostat. The switching times were recorded with a storage-scope.

3. Bistable properties of the thin CdS film

In order to achieve all-OB one needs a nonlinear interaction between the exciting source and the illuminated medium. In our case all-OB arises from the nonlinear dependence of the transmittance on temperature $Tr(T)$ shown in Fig. 1. The measurement of $Tr(T)$ was performed by heating the sample continuously under excitation with the focused cw 514.5 nm

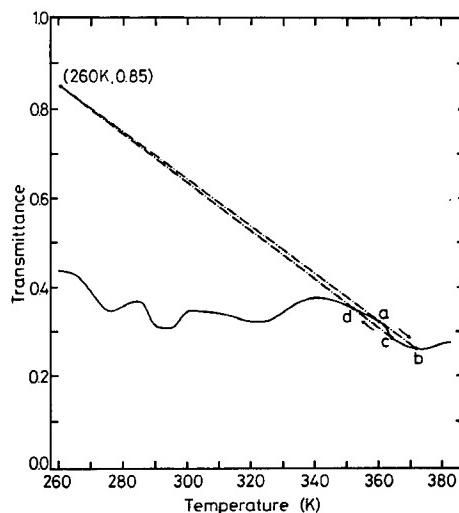


Fig. 1. The transmission dependence on temperature of the thin CdS : Cu film. The measurement was performed with the focused cw 514.5 nm Ar^+ -laser line of 6 mW. The drawn tangents on the measured transmittance starting at (260 K, 0.15) represent the graphical solution of the bistability in transmission. The decrease and increase in the transmittance during the switch down (a → b) and up, respectively is about 20%.

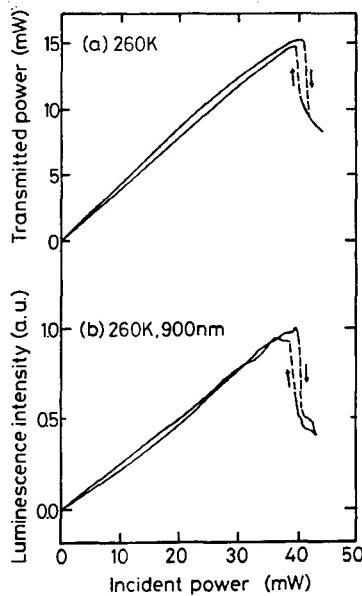


Fig. 2. (a) All-optical bistability of the transmitted power of the CdS : Cu film. (b) All-optical bistability of the luminescent emission at 900 nm of the thin CdS : Cu film. Both measurements were performed with the 514.5 nm line at an ambient temperature of 260 K.

line of 6 mW. The graphical solution for all-OBT is defined by the two drawn tangents on $Tr(T)$ starting at the so-called operating point (OP) which has the co-ordinates (T_a , $1-Re$) where T_a is the ambient temperature of the film (260 K) and Re the reflectance (0.15) of the sample [4, 6, 7]. Precisely, one must take into account that Re depends on temperature and the co-ordinates of OP shift during the bistable switch [7]. However, by measuring coincidentally transmittance and reflectance of the thin CdS film, we observed that even if a bistable loop in the transmitted power has been measured no bistability in the reflectance has taken place. This phenomenon of the decoupling between Tr and Re of the thin CdS film under high excitation is at first difficult to understand and discussed in detail elsewhere [7, 8]. In the current context it is reasonable to assume a constant OP. The indicated points on $Tr(T)$ in Fig. 1 define the bistable jump

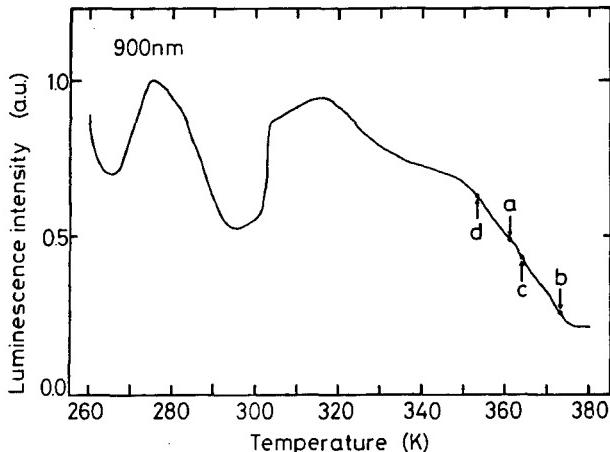


Fig. 3. The dependences of the luminescence intensity on temperature of the thin CdS : Cu film at 900 nm. The sample was excited with the focused cw 514.5 nm Ar⁺-laser line of 6 mW. The indicated points correspond with those of Fig. 1.

down ($a \rightarrow b$) and up ($c \rightarrow d$). Figure 2(a) reveals the result for all-OBT with a contrast of 38% which is larger as the expected value of 20% from Fig. 1. The deviation probably arise from the peculiar shape of $Tr(T)$ which indicates a diffuse gap of the thin CdS film and the excited carriers might influence the band structure in a considerable manner which cannot be deduced from Fig. 1. Figure 2(b) shows the all-OBL at 900 nm. The measured hysteresis occurs as the all-OBT in a clockwise fashion since $I_{Lum}(T)$ decreases in the considered temperature range as can be seen in Fig. 3. Furthermore, by plotting points in Fig. 3 at the corresponding T values of Fig. 1, one finds a quite good agreement between the measured contrast [40% down and 30% up in Fig. 2(b)] and the expected one [47% ($a \rightarrow b$) and 30% ($c \rightarrow d$) in Fig. 3].

4. Switching behaviors

Figures 4(a) and 4(b) show the switches of all-OBT and all-OBL, respectively. Clearly, at the threshold of the bistable switch, I_{Lum} contains a faster decay over time (slope: -23% div.) which is not observed in the transmitted power (slope: -10% div.). We point out that the luminescent response of the sample was measured with the monochromator adjusted exactly in the transmitted beam and no modifications of the foci used have been performed during the measurements of Figs 4(a) and 4(b). Therefore, we have detected the luminescence signal which was in the transmitted beam included. As a consequence, the thermal relaxation time is exactly the same for both measurements. We therefore conclude that during the switch of all-OBL both photo-thermally induced and electronic effects take place, i.e., all-OBL is a *photo-thermally induced electronic bistability*. The statement is supported by the observation that at various wavelengths

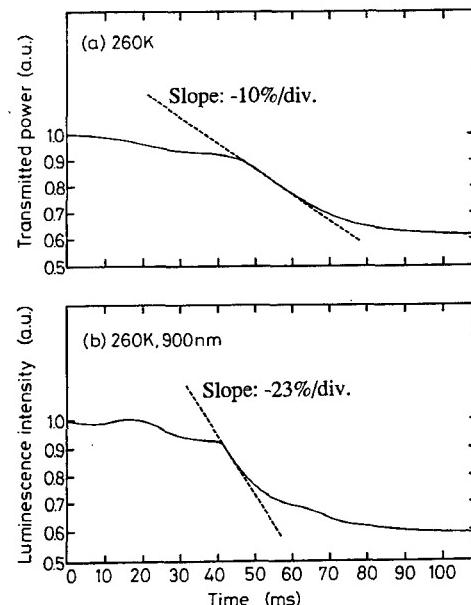


Fig. 4. Switch-off characteristics of the bistabilities in (a) transmission and (b) luminescence at 260 K. The measurements were performed with the same experimental conditions. Clearly, the switch of the luminescence contains a faster component as indicated by the broken lines (slope: -23%/div.) than the switch of the transmitted power (slope: -10%/div.).

different decays of all-OBL were observed. The time scale (≤ 20 ms) for the switching process of all-OBT in Fig. 4(a) corresponds quite reasonably with the expected thermal relaxation time for the focus used [3]. It must be noted, that the complete bistable switch of I_{Lum} cannot be faster than the thermal relaxation time since, as pointed out above, the initial and final states of all-OBL are determined by $I_{\text{Lum}}(T)$. However, if electronic (faster) processes take place between the initial and final states, partly the switch occurs faster than in a dominantly thermally driven process as shown by Figs 4(a) and 4(b).

5. Conclusion

For the first time, we have shown the differences between the switching behaviors of all-OBT and all-OBL. On the basis of the presented data in this paper, it appears that at the beginning of the switch of all-OBL – in the moment when T starts to increase – I_{Lum} decreases due to the radiative recombination lifetime and not due a thermal relaxation process. The results encourage to apply radiative transitions with short lifetimes ($< 10^{-9}$ s) to all-OBL in order to overcome the restriction of the technical applica-

tions of common photo-thermally induced all-OBs due to the rather slow thermal relaxation times.

Acknowledgements

We are thankful to H. Nguyen Cong and Prof. P. Chartier for the excellent sample preparation. One of us (B.U.) acknowledges the financial support by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under the project JO 824-PHY.

References

1. Kobayashi, T., IEICE Trans. Fundamentals, **E75-A** (1992) 38.
2. Ullrich, B., Kazlauskas, A., Zerlauth, S., Nguyen Cong H. and Chartier, P., "Extended Abstracts of the 1993 International Conference on Solid State Devices and Materials" (Makuhari 1993), p. 669.
3. Ullrich, B., Kazlauskas, A., Zerlauth, S. and Kobayashi, T.: J. Crystal Growth **138**, 234 (1994).
4. Ullrich, B., Bouchenaki, C. and Roth, S., Appl. Phys. **A53**, 539 (1991).
5. Bouchenaki, C., Ullrich, B., Zielinger, J. P., Nguyen Cong H. and Chartier, P., J. Opt. Soc. Am. **B8**, 691 (1991).
6. Ullrich, B., and Bouchenaki, C., Jpn. J. Appl. Phys. **30**, L1285 (1991).
7. Ullrich, B., and Zerlauth, S., Nonlinear Opt. **7**, 123 (1994).
8. Ullrich, B., Nonlinear Opt. **4**, 63 (1993).

Selective-Area MOVPE for InP-Based Optoelectronic Components

P. Tidemand-Petersson, O. Albrektsen and J. Salzman*

Tele Danmark Research, Lyngsø Allé 2, DK-2970 Hørsholm, Denmark

Received May 4, 1994; accepted June 15, 1994

Abstract

Selective-area MOVPE of InP-based semiconductor material is discussed. As an example of its applicability, the use of the technique to make distributed feedback gratings with sub-micron period and modulated coupling coefficient κ is described. The grating pattern and the lateral variation in growth rate of the InGaAs(P) material are defined and controlled by a SiO₂ mask layer.

1. Introduction

Selective-area metalorganic vapour phase epitaxy (SA-MOVPE) has gained increased interest as a fabrication tool for optoelectronic components during the last years, see e.g. Refs [1–4]. This is to a large extent based on the applicability of the technique to make optoelectronic integrated circuits (OEICs).

However, before using SA-MOVPE in the fabrication of components, the technique must evidently be mastered to a degree where monocrystalline planar semiconductor material of well-defined composition can be grown on a masked surface. This makes a thorough study of the growth dependence on various parameters a necessary prerequisite for the application. In this paper the investigation of growth properties and an application of SA-MOVPE: the fabrication of gratings for distributed feedback (DFB) laser diodes, are described.

2. Selective-area MOVPE

The most common way to achieve SA-MOVPE is to deposit a thin dielectric layer on the semiconductor surface, to pattern this layer in order to expose the desired areas for selective growth, and finally to perform the MOVPE growth.

An important finding in earlier studies of this technique is that the epitaxial growth in the exposed areas proceeds at an increased rate, with the growth enhancement depending on the relative coverage of the masked areas. In the literature the reported specific dependence of the growth enhancement on the mask geometry varies considerably. The reason for this discrepancy is that growth rates in selec-

tive epitaxy are governed by the characteristic lifetimes of the reacting species (gas diffusion, adatom diffusion on the mask, adsorption on the growth surface, reaction and desorption lifetimes). These lifetimes depend on the growth parameters such as pressure, temperature, V/III ratio, nominal growth rate, etc. Hence, the earlier reports could only provide data on growth enhancement for the particular growth parameters and mask geometry used.

In the present studies, the strategy was first to optimize the growth morphology, with special emphasis on layer planarity and smoothness of interfaces and sidewalls. This was done by growing InP with thin InGaAs marker layers. All growths were performed on exact [100] oriented InP substrates in a horizontal reactor at 20 mbar, using trimethylgallium (TMGa) and trimethylindium (TMIIn) as group III sources and PH₃ and AsH₃ as group V sources. The growth temperature was typically in the range 640–670 °C, and the growth rate was varied from 0.3–2.5 μm/h. The non-optimized growths showed e.g. interrupted and non-planar InGaAs layers and rough [111] sidewalls when growing in 2–5 μm wide openings in SiO₂-masks. After optimization, which included a decrease in growth rate and an increase in temperature and V/III-ratio, an optimized structure with smooth, planar and buried marker layers and flat sidewalls was obtained, as can be seen from Fig. 1.

With the growth parameters yielding good morphology, the growth-rate enhancement, due to the partial cover of the

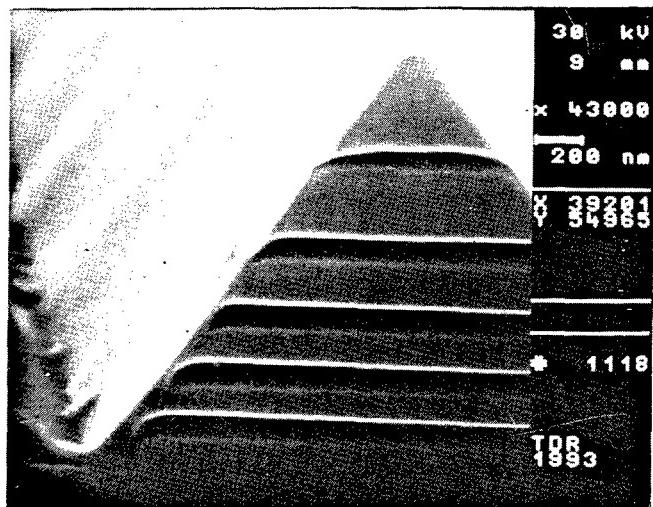


Fig. 1. Scanning electron micrograph showing a nearly optimal structure of InP with InGaAs marker layers grown by selective-area MOVPE.

* Permanent address: Dept. of Electr. Eng. and Solid State Inst., Technion, Israel Inst. of Technology, Haifa 3200, Israel.

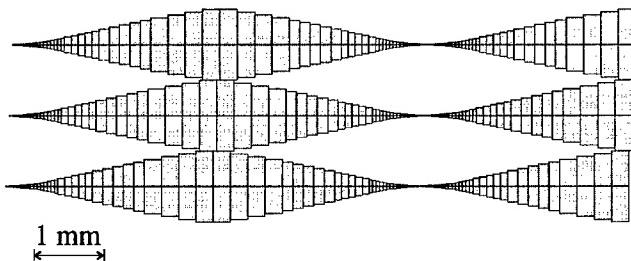


Fig. 2. Schematic view of the calibration mask used in the studies of growth properties. During growth the shaded area was covered with dielectric material. The width of this area varies from 9 μm to 1000 μm .

substrate with an SiO_2 layer, was studied with a specially designed calibration pattern, shown in Fig. 2. Having in mind various photonic devices, we studied the growth in a narrow opening along the axis of the "near-rhombic" mask. In that way the growth conditions could be evaluated as a function of the width of the surrounding mask. The opening was either a $\sim 3 \mu\text{m}$ wide "waveguide" or a series of 20 μm long grating lines perpendicular to the axis. The period of the gratings was about 235 nm.

The local epitaxial layer thickness at different points along the SiO_2 pattern was measured with a surface profiler (Dektak) or with a scanning electron microscope (SEM). As an example, the waveguide thickness (measured with the Dektak) is shown in Fig. 3 as a function of the width (on each side) of the surrounding SiO_2 mask. In this case nominally 500 nm InP was grown, so even where the mask is very narrow (3 μm on each side of the waveguide) a slightly enhanced growth rate is observed. The waveguide of this growth was seen to change from a simple triangular shape, like the one shown in Fig. 1, to the more complicated shape shown in Fig. 4, in which a growth over the SiO_2 mask is seen to have taken place. The transition (see Fig. 3) from the fast increase in thickness for small values of the mask width, to the more moderate increase observed for larger mask width values is presumably due to this change in shape. This assumption is under further investigation.

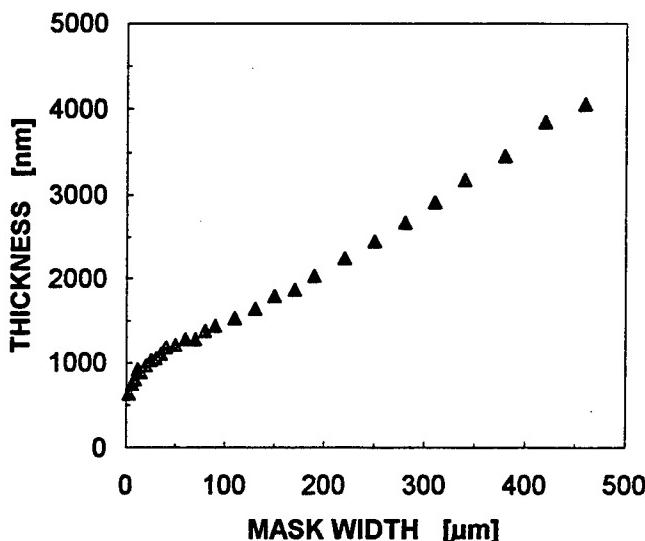


Fig. 3. Measured thickness of a narrow waveguide shown as a function of the width of the surrounding SiO_2 mask. For small values of the mask width the grown waveguide was observed to be of regular triangular shape, while for larger values the shape was as shown in Fig. 4.

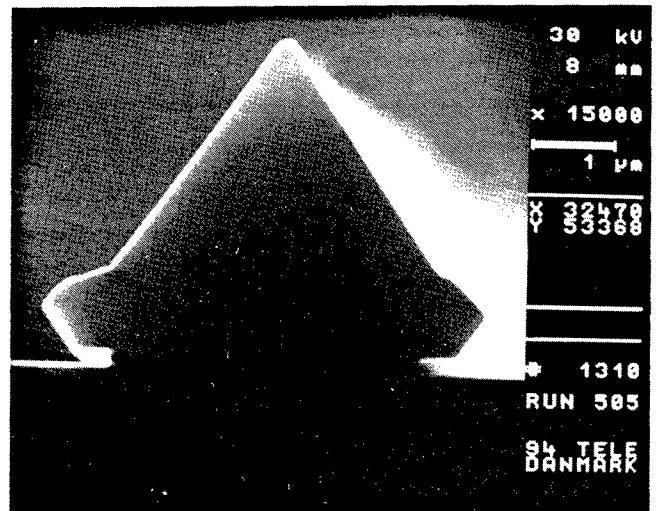


Fig. 4. Scanning electron micrograph showing the shape of a waveguide grown in the opening along the axis of the calibration mask shown in Fig. 2. The large width of the mask is seen to have caused the waveguide to broaden considerably compared to the opening in the SiO_2 .

3. Grown DFB gratings

As an example of the applicability of SA-MOVPE to optoelectronic devices, we have used the technique to make DFB gratings. The motivation for this work is to enable fabrication of gratings with coupling coefficient κ varying along the laser cavity. This is important for the design of a laser cavity to yield nearly uniform intensity distribution along its axis, and thus preventing spatial hole burning, see e.g. Refs [5–7].

The concept for growing DFB gratings is shown schematically in Fig. 5: In a first growth a "normal" laser structure consisting of an n-InP buffer, an active region with confinement layers and quantum wells, and a 50–250 nm thick InP spacer layer is made. Then a 35 nm thick SiO_2 mask layer is deposited by plasma-enhanced CVD and patterned by conventional photolithography to define the variation of growth rate in the subsequent selective-area growth. In this case the 3 μm wide opening in the calibration mask is replaced by a 235-nm period grating, consisting of a regular pattern of 20 μm long lines written by E-beam lithography and transferred to the mask layer by CHF_3 reactive ion etching (RIE). An InGaAs(P) grating is then grown by SA-MOVPE in the exposed areas between the SiO_2 -lines. After removal of the remaining SiO_2 , the p-InP cladding layer and a p⁺-InGaAs contact layer are regrown to complete the laser structure. The latter step is facilitated by terminating the selective growth by a thin InP cap layer. By this concept DFB gratings can be made without any etching of semiconductor material.

A variation of the width of the surrounding SiO_2 mask, implies also in this case a variation of the growth rate and, hence, the thickness of the grown grating lines. This principle is depicted in Fig. 6.

4. Results and discussion

Grown gratings were made according to the outlined procedure. Performing growth at 640 °C and a nominal growth rate of 0.3 μm per hour, we obtained grating lines with smooth sidewalls. They were observed to exhibit better mor-

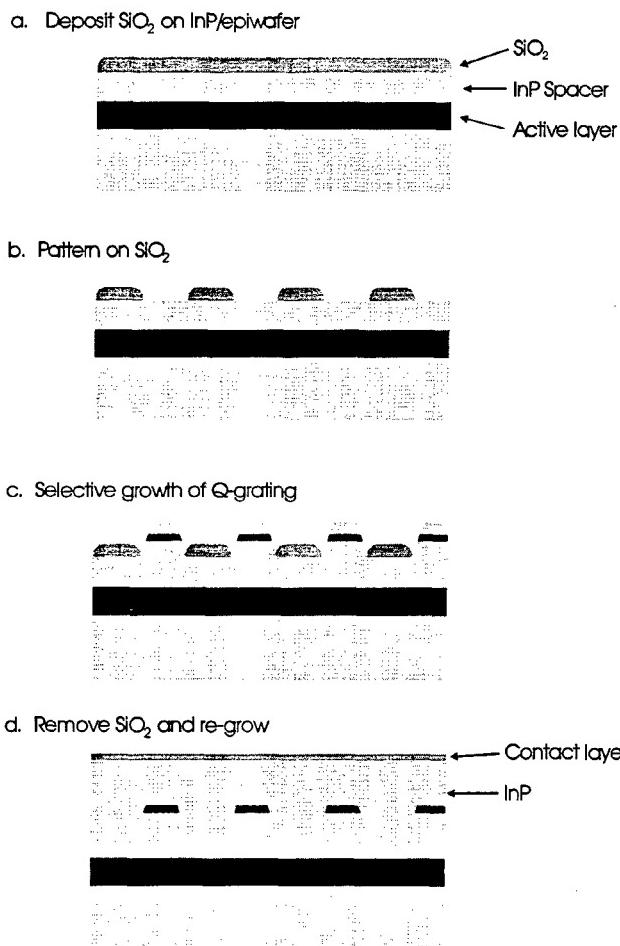


Fig. 5. Fabrication procedure for selectively grown DFB gratings.

phology (smoother lines) than the defining pattern in the SiO_2 layer, presumably because of a self-smoothing effect of the slow-growth [111] facets.

An intermediate state in the fabrication – corresponding to Fig. 5(c) – is shown in Fig. 7. Here a nominally 25 nm thick layer was grown by SA-MOVPE. Assuming a spacer layer of fixed thickness, the observed variation in grating layer thickness corresponds to a change in the coupling coefficient by a factor of 6. The growth of considerably thicker grating layers resulted in partly merged neighbouring grating lines in the highly enhanced regions, with the maximum applicable thickness also determined by the width

of the opening between the SiO_2 grating lines. This line-width depends mainly on the exposure time in the E-beam lithography.

The gratings shown in Fig. 7 were grown directly on an InP substrate, whereas in Fig. 8 is shown a grating grown on a real laser structure, as can be seen from the presence of 3 quantum wells embedded in an InGaAsP confinement structure. The selective growth step consisted here of nominally 3 nm InP, 19 nm InGaAs and 3 nm InP. This structure was subsequently buried – corresponding to Fig. 5(d). Figure 8 reveals a very uniform grating. The resulting coupling coefficient depends not only on the thickness of the

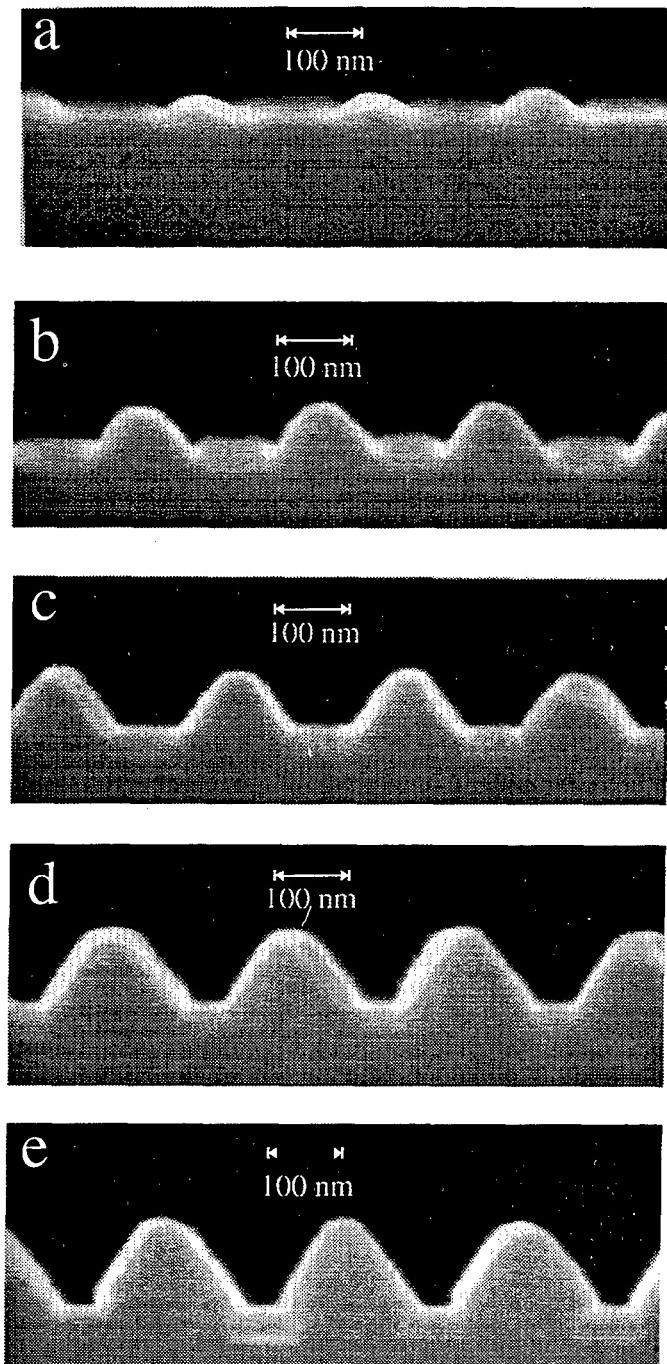


Fig. 7. SEM cross sections of gratings grown selectively along the axis of the calibration mask. The photographs were taken at 0.5 mm intervals. The height of the lines varies from 25 nm (a) to 160 nm (e). The rectangular parts in (a) are the defining SiO_2 stripes. They are also visible – almost buried – in (e).

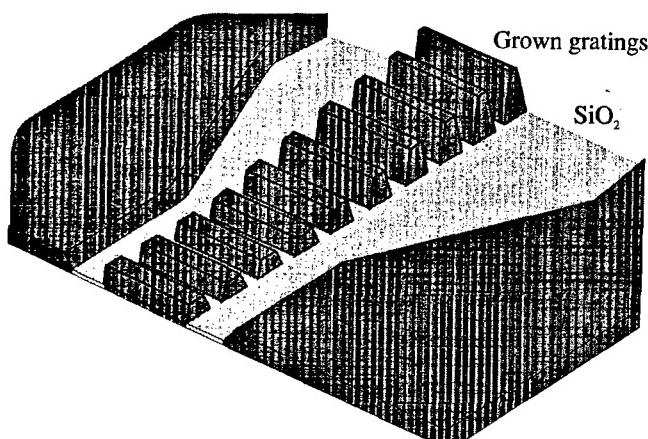


Fig. 6. Schematic illustration of the method of growing DFB gratings with varying thickness of the grating lines: a broader surrounding SiO_2 mask implies a thicker grating line.

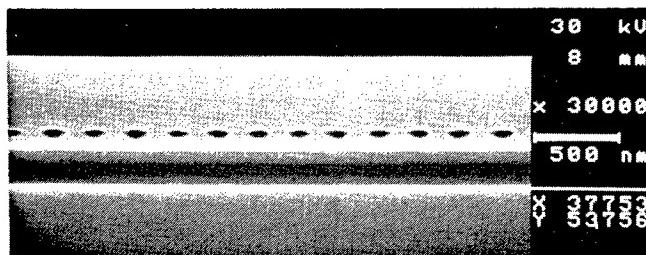


Fig. 8. SEM cross section of a buried InGaAs grating grown selectively on top of the active layers of a laser structure with 3 quantum wells.

grating layer but, naturally, also on the distance from the active laser region, i.e. on the thickness of the spacer layer. For a 200 nm thick spacer layer we have estimated (using Ref. [8]) that a variation in corrugation depth from 25 nm to 150 nm corresponds to a variation in coupling coefficient from 40 cm^{-1} to 250 cm^{-1} .

5. Conclusion

Selective-area MOVPE was performed, and the growth parameters optimized to obtain good morphology. The

dependence of growth enhancement on mask geometry was studied under these growth conditions. Finally, a new method for making gratings by selective-area MOVPE was presented. Using the prescription proposed here, DFB gratings can be made solely by growth, i.e. without involving any etching of semiconductor material. The method can be used to achieve gratings with controllable axial variations in the coupling coefficient in a reproducible way.

References

1. Galeuchet, Y. and Roentgen, P., *J. Cryst. Growth* **107**, 147 (1991).
2. Caneau, C. *et al.*, *J. Cryst. Growth* **124**, 243 (1992).
3. Thrush, E. J. *et al.*, *J. Cryst. Growth* **124**, 249 (1992).
4. Sasaki, T., Kitamura, M. and Mito, I., *J. Cryst. Growth* **132**, 435 (1993).
5. Schranks, T. and Yariv, A., *Appl. Phys. Lett.* **56**, 1526 (1990).
6. Morthier, G., David, K., Vankwikelberge, P. and Baets, R., *IEEE Photon. Technol. Lett.* **2**, 388 (1990).
7. Talneau, A., Charil, J., Ougazzaden, A. and Bouley, J. C., *Electron. Lett.* **28**, 1395 (1992).
8. Jonsson, B., "Field – A Versatile Program for Modeling One-Dimensional Optical Waveguides" (Chalmers University of Technology, Göteborg 1992), Technical report No. TR 92316.

Interdiffusion and Phase Formation During Thermal Processing of Co/Ti/Si(100) Structures

J. Cardenas, S. Hatzikonstantinidou, S.-L. Zhang, B. G. Svensson and C. S. Petersson

Royal Institute of Technology, Solid State Electronics, Electrum 229, S-164 40 Kista, Stockholm, Sweden

Received June 12, 1994; accepted June 15, 1994

Abstract

Rapid thermal processing of Co/Ti/Si(100) structures has been investigated over a wide temperature range, from 300°C to 1100°C. Titanium and cobalt layers, 10 nm and 20 nm thick respectively, were consecutively deposited on Si(100) substrates using an e-beam evaporator. X-ray diffraction, Secondary Ion Mass Spectrometry and Rutherford Backscattering Spectrometry were employed to analyse interdiffusion and phase formations. The present results revealed the formation of CoSi₂ when annealing at 750°C. The CoSi₂ was formed epitaxially and yielded a minimum channelling yield of approximately 8.5% for layers annealed at 1100°C. Evidence for the presence of a CoSi layer, when annealing between 650°C and 800°C, is presented.

1. Introduction

CoSi₂ is a promising candidate for metallization in future Si integration technology mainly because of its low resistivity and the possibility of epitaxial growth on Si because of its small lattice mismatch with respect to Si [1, 2]. It has recently been reported [3] that epitaxially grown CoSi₂, once formed, is thermally stable up to 1100°C. Thus it is of technological interest to investigate the techniques for CoSi₂ epitaxial growth on Si(100), since such (100) oriented wafers are most frequently used in device fabrication. The epitaxial growth of CoSi₂ on Si(100) has in general been achieved with elaborate techniques such as ion-beam synthesis [4], deposition under ultra high vacuum conditions on an atomically clean surface [5], or by hot co-deposition of Co and Si followed by a Si cap layer [6]. One alternative and attractive approach was recently presented by Wei *et al.* [7] who obtained epitaxial CoSi₂ layers on Si(100) substrates by heat treating a Co/Ti/Si(100) structure. The first results on epitaxial growth of CoSi₂ on Si(100) substrates with Co/Ti/Si(100) structures were actually published by Yang and Bene [8], but Wei *et al.* [7] were the first ones to attribute the epitaxial growth to the intermediate Ti layer. They observed, when annealing Co/Ti/Si(100) structures by rapid thermal processing (RTP), an inversion between the Ti and the Co layers which was accompanied by the epitaxial CoSi₂ formation. Two arguments have been proposed which may to some extent account for such a formation. The first one is the ability of Ti to getter oxygen [7], leaving a clean Si surface; the second one is the reduced supply of Co atoms to the Si surface in the presence of a intermediate Ti-layer [9]. However, the fundamental mechanisms that lead to the formation of epitaxial CoSi₂ in the presence of Ti remain obscure. The present work is intended to elucidate the process of Co-Ti-Si inter-diffusion and the sequence of

phase formation leading to the growth of the epitaxial CoSi₂ layer.

2. Experimental procedure

Ti and Co layers, 10 nm and 20 nm thick, respectively, were consecutively deposited, using a dual source e-beam evaporator, on Si(100) wafers of p-type (75 mm in diameter) forming a Co/Ti/Si(100) structure. The base pressure in the evaporator was 2×10^{-8} Torr. Annealing of the deposited Co/Ti/Si structures was performed, with an AG Associates Heatpulse RTP system, in N₂. For the first set of wafers, the annealing temperature extended from 300°C to 850°C. For the second set of wafers the temperature ranged from 950°C to 1100°C. The duration of the annealing was 30 s for the first set of wafers and 10 s for the second set. The temperature was monitored with a thermocouple below 600°C, and with a pyrometer above 600°C.

The samples were analysed by X-ray Diffraction (XRD), Secondary Ion Mass Spectrometry (SIMS) and Rutherford Backscattering Spectrometry (RBS). The XRD measurements were performed using a Θ–2Θ X-ray diffractometer with Cu radiation and a postsample monochromator to identify the formed phases. The RBS analysis, utilising ⁴He ions with an acceleration energy of 2 MeV, was used in random mode to determine the stoichiometry of the layers formed and in channelling mode to characterise the epitaxial growth of CoSi₂. In order to facilitate the RBS channelling study, a one step or a two step etching procedure was used to selectively remove the upper layer(s) formed on the CoSi₂ layer. The one step procedure used a H₂O₂ : H₂SO₄ (1 : 3) solution, and the two step procedure consisted of a H₂O : H₂O₂ : NH₃ (1 : 1 : 5) solution followed by the H₂O₂ : H₂SO₄ (1 : 3) solution. A Cameca IMS 4f micro-analyser was used for the SIMS analysis. A primary sputtering beam of ¹³³CS⁺ ions, with an energy of 4.0 keV, was rastered over an area with a size of approximately 100 × 100 μm², and secondary ions were collected from the central part (diameter ~8 μm) of the sputtered crater. Secondary ions of the type MCs⁺ were detected, where M denotes the matrix or impurity species to be monitored. It has recently been demonstrated by different authors that this so called “MCs⁺SIMS” significantly reduces the matrix ionisation effects, compared to what is obtained with atomic ions and is suitable for quantitative analysis of major elements [10, 11]. “MCs⁺SIMS” has been employed in the present study to reveal the compositions of the different

layers formed and not primarily to monitor small concentrations of impurities. Furthermore, sheet resistance measurements were carried out in order to help monitoring the silicide formation.

3. Results and discussion

The as-deposited Ti and Co layers, both hexagonal close packed, were found to grow with the preferential orientation (002), which is in agreement with results reported previously in the literature [12]. After annealing at 300 °C some intermixing of Co and Ti occurs, as revealed by the SIMS spectra in Fig. 1(A). A considerable amount of Co atoms was found at the Ti/Si interface, suggesting Co diffusion at

300 °C, however most of the Co layer remained intact. Oxygen was found to be present both at the Co/Ti and Ti/Si interfaces. The SIMS results suggest a much higher Co concentration in the Ti region adjacent to the Ti/Si interface, than what is expected according to the Co-Ti phase diagram [13]. This may indicate a silicide formation at the interface presumably involving Si, Ti and Co. The XRD analysis confirms (Fig. 2) the continuing presence of elemental Co and Ti phases after annealing at this temperature. However, no compound formation was observed; the amount of silicide was probably too small to be detected.

Annealing at 400 °C resulted in a broad XRD peak appearing around 2Θ equal to 37.5° (Fig. 2) which can be assigned to a very thin and highly distorted Ti layer.

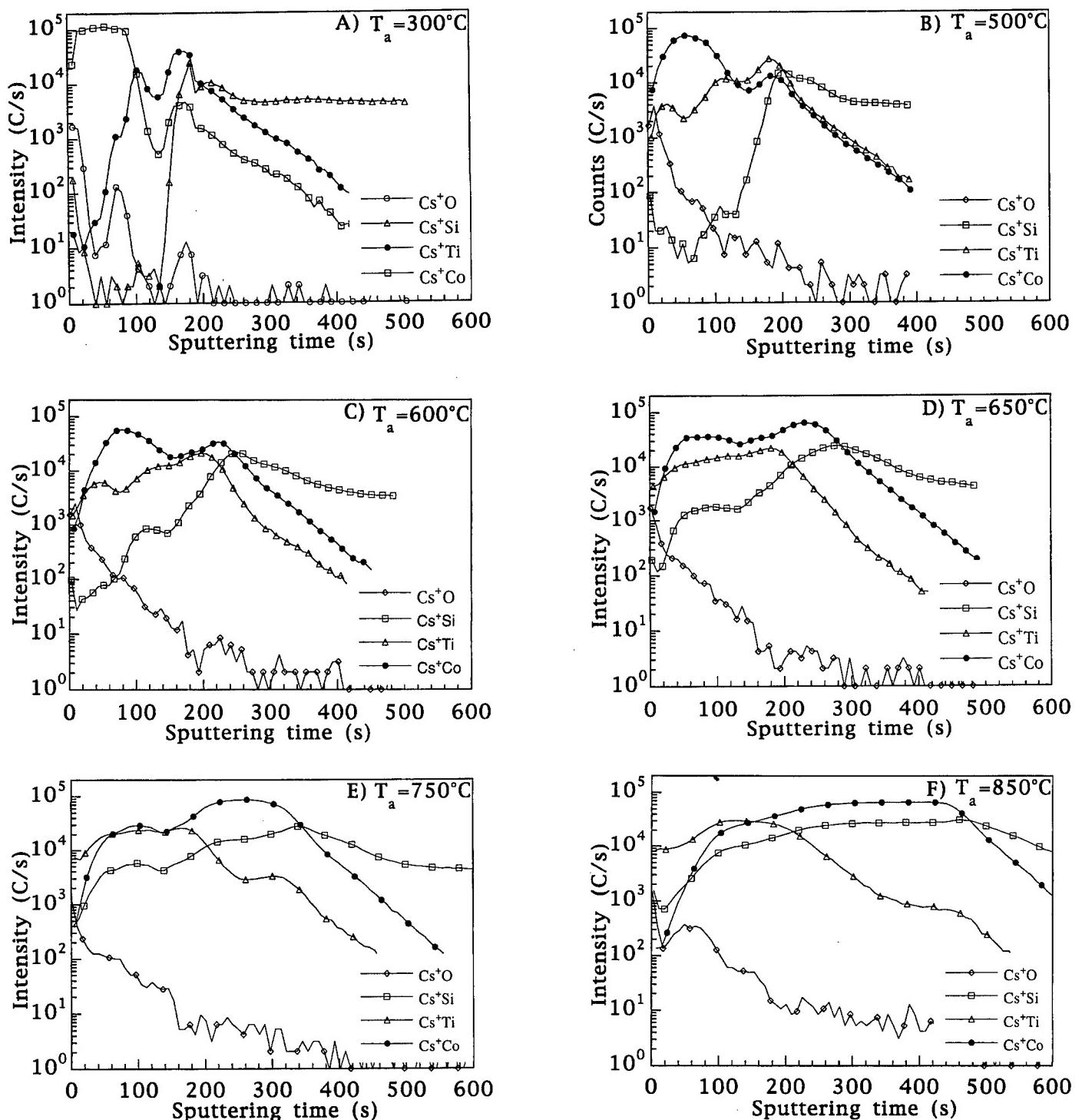


Fig. 1. SIMS results showing depth profiles of the Co(20 nm)/Ti(10 nm)/Si(100) structures annealed at different temperatures.

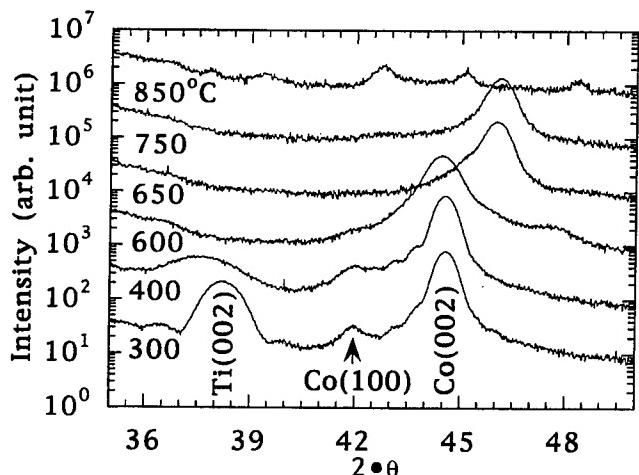


Fig. 2. XRD patterns of Co(20 nm)/Ti(10 nm)/Si(100) structures annealed for 30 s at the temperatures indicated.

Annealing above 400 °C resulted in a considerable mixing of the Ti containing layer with Co atoms so that the Ti layer could not longer be differentiated as a distinguishable phase. SIMS data showed the presence of Ti at the surface [Fig. 1(b)]. It is noteworthy that the SIMS measurements revealed an increased Ti to Co ratio in the near vicinity of the sample surface, i.e. on top of the Co layer, by approximately two orders of magnitude after the annealing temperature was raised from 400 °C to 500 °C. The ratio stayed approximately constant for samples annealed between 500 °C and 600 °C. The surface Ti was most probably in the form of TiN(O) [14], between 500 °C and 600 °C.

After annealing between 600 °C and 650 °C there was a drastic change in the XRD spectra (Fig. 2) and in the composition profiles obtained by SIMS [Figs 1(C) and (D)] as well. The Ti/Co ratio formed a distinct plateau in the near vicinity of the surface, indicating a dissolution of the original Co layer. This transformation was correlated with the abrupt change in the XRD spectra, manifested by the disappearance of the peak attributed to Co and the appearance of a peak with 2θ around 46° (Fig. 2). After annealing between 650 °C and 800 °C, this peak, not yet assigned, gradually decreased with temperature and vanished at 850 °C. The SIMS data revealed a constant Co to Si ratio, adjacent to the substrate, in samples annealed in a temperature range between 650 °C and 750 °C. After annealing, between 750 °C and 850 °C, the XRD results showed a gradual increase of the two diffraction peaks at approximately 33.8° [Fig. 3(A)] and 71° [Fig. 3(B)]. These two peaks are from the (200) and (400) diffraction planes of the CoSi₂ phase. The simultaneous appearance of these two diffraction peaks may exclude the recently proposed assignment of the 71° peak to the CoSi phase [15].

The ratio between Co and Si stayed constant in a broad layer close to the Si substrate; it corresponds to the CoSi₂ layer in the sample annealed at 850 °C [Fig. 1(F) and 2]. This Co to Si ratio was used in an attempt to identify the possible silicide phase formed between 650 °C [Fig. 1(D)] and 750 °C [Fig. 1(E)]. Calculations based on the relative changes in the Co to Si ratio, close to the Si substrate, imply strongly that a CoSi layer was present in samples annealed between 650 °C [Fig. 1(D)] and 750 °C [Fig. 1(E)]. This CoSi layer may be correlated with the diffraction peak appearing around 46° for the samples annealed between

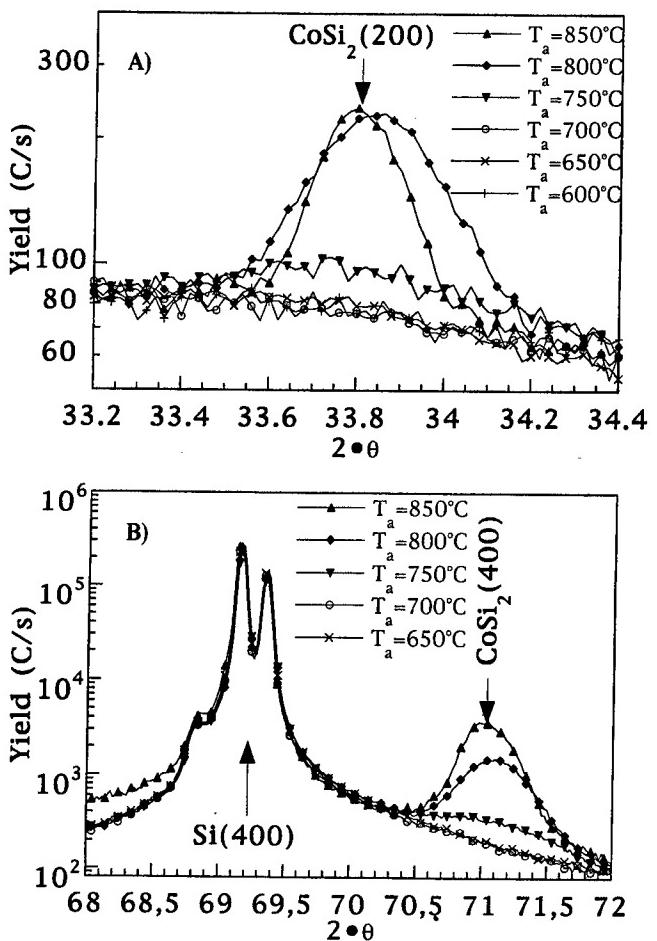


Fig. 3. XRD spectra showing the diffraction peaks from the CoSi₂ A) (200) and B) (400), for Co(20 nm)/Ti(10 nm)/Si(100) structures annealed at various temperatures.

650 °C and 800 °C. None of the small diffraction peaks in the 2θ range between 35° and 50° matches any cobalt silicide for the sample annealed at 850 °C (Fig. 2). On the other hand, these peaks match well with titanium nitride. The absence of any peaks other than the CoSi₂ (200) and (400) indicates the epitaxial alignment of the CoSi₂ layer.

RBS results confirmed the epitaxial growth of CoSi₂ on the Si(100) substrate (Fig. 4). The minimum yields in the channeling spectra of the CoSi₂ layers formed at 850 °C, 950 °C and 1100 °C were approximately 18%, 13% and 9%,

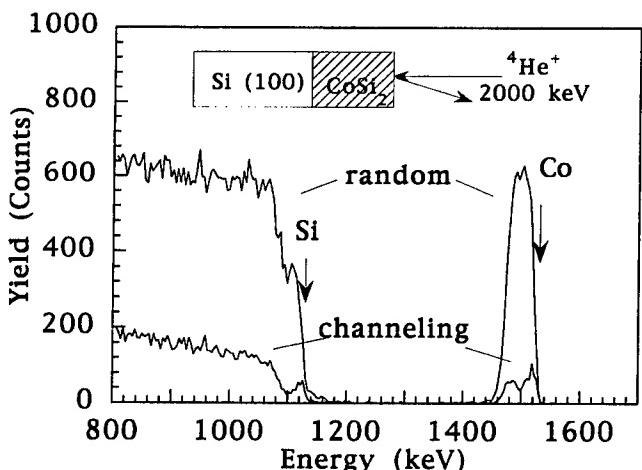


Fig. 4. RBS random and channeling spectra of the Co(20 nm)/Ti(10 nm)/Si(100) structure, annealed at 1100 °C, displaying a minimum channeling yield of 8.5%. The sample was exposed to the two step etching procedure prior to the measurement.

respectively. These values are comparable to those reported by other authors [14]. Above 850 °C, the total thickness of the epitaxial CoSi₂ layer depended only weakly on the annealing temperature. The data on unetched samples revealed a complex top surface layer involving Co, Ti, Si and probably N.

Although the sequence Co₂Si, CoSi and CoSi₂ has been repeatedly reported for studies relating the furnace annealing of Co-Si samples, we have not conclusively identified the presence of Co₂Si in the course of the present investigation.

4. Conclusions

We have shown that the Co layer in a Co/Ti/Si(100) structure ceased to exist when annealing above 650 °C. The SIMS data, for samples annealed between 650 °C and 750 °C indicated strongly that a CoSi layer formed near the Si substrate. Between 750 °C and 850 °C XRD results showed conclusively a gradual formation of an epitaxial CoSi₂ layer. RBS channelling measurements gave minimum yields of 18%, 13% and 9%, after annealing at 850 °C, 950 °C and 1100 °C, respectively.

Acknowledgements

Fruitful discussions with Prof. F. M. d'Heurle are gratefully appreciated. The RBS measurements were kindly performed by J. Saarilahti, VTT,

Finland. Financial support was partly received from the Swedish Board for Technical Development (NUTEK).

References

- Hensel, J. C., Tung, R. T. and Poate, J. M., *Appl. Phys. Lett.* **44**, 913 (1984).
- Nicolet, M.-A. and Lau, S. S., in "VLSI Electronics Microstructure Science" (Edited by N. G. Einspruch and G. B. Larrabee) (Academic, New York 1983), vol. 6, p. 402.
- Hsia, S. L., Tan, T. Y., Smith, P. and McGuire, G. E., *J. Appl. Phys.* **70**, 7579 (1991).
- Maex, K., Vanhellemont, J., Petterson, S. and Lauwers, A., *Appl. Surf. Sci.* **53**, 273 (1991).
- Yalisove, S. M., Tung, R. T. and Batstone, J. L., *Mat. Res. Symp. Proc.* **116**, 439 (1988).
- Mantl, S. and Bay, H. L., *Appl. Phys. Lett.* **61**, 267 (1992).
- Wei, C.-S., Fraser, D. B., Dass, M. L. A. and Brat, T., in VI International IEEE VLSI Multilevel Interconnection Conference, Santa Clara, CA, (1989), p. 241.
- Yang, H.-Y. and Bene, R. W., *J. Appl. Phys.* **59**, 1525 (1986).
- Setton, M., and Van der Spiegel, J., *Appl. Surf. Sci.* **38**, 62 (1989).
- Gao, Y., *J. Appl. Phys.* **64**, 3760 (1988).
- Magee, C. W., Harrington, W. L. and Botnick, E. M., *Int. Mass. Spectrom. Ion Proc.* **103**, 45 (1990).
- Barmak, K. et al., *Mat. Res. Symp. Proc.* **238**, 575 (1991).
- Murray, J. L., "Binary Alloy Phase Diagrams" (Edited by T. B. Massalski) (ASM, Metals Park, Ohio 1986), vol. I, p. 803.
- Liu, P. et al., *J. Appl. Phys.* **74**, 1700 (1993).
- Hong, F., Rozgonyi, G. A. and Patnaik, B. K., *Appl. Phys. Lett.* **64**, 2241 (1994).

Laser Ablation Deposition as a Preparation Method for Electronic Materials

S. Leppävuori

Microelectronics and Material Physics Laboratories, University of Oulu, P.O. Box 400, FIN-90571 Oulu, Finland

Received May 26, 1994; accepted June 15, 1994

Abstract

In the development of new electronic materials and devices, laser ablation deposition offers a quick and relatively easy technique to realise complicated material compositions and multilayer structures. This paper presents laser ablation deposition of different types of thin film materials, each having potential use in electronic applications. The materials are: $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ superconducting films and $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{PrBa}_2\text{O}_{7-\delta}$ superlattices; piezoelectric lead zirconium titanate (PZT) and PZT/ $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ heterostructures; Cu(In,Ga)Se₂ semiconductor photovoltaic materials and amorphous diamond-like carbon films. The deposition of these materials was carried out mostly using a pulsed XeCl excimer laser, either in an in situ process on a heated substrate or by post deposition annealing.

1. Introduction

The deposition of high temperature superconductors has shown laser ablation deposition (LAD) to be a simple and versatile thin film preparation method. Especially in the development of new materials, LAD offers a quick and relatively easy technique to realise complicated material compositions and multilayer structures. The main advantages of the method are the reproducibility of the target composition, the possibility to deposit a wide variety of materials and layered structures and a quick and relatively simple experimental arrangement.

In this paper, pulsed laser ablation deposition of different types of thin film materials made in the University of Oulu, is presented [1–5]. The materials are very different in their deposition properties offering a challenge to any method: $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ superconducting films and $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{PrBa}_2\text{O}_{7-\delta}$ superlattices; doped-lead zirconium titanate (PZT), a multinary ferroelectric oxide, and PZT/ $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ heterostructures; Cu(In, Ga)Se₂ a ternary or quaternary compound semiconductor consisting of more and less volatile elements and amorphous diamond-like carbon, ablated from a high-melting point element.

2. Laser ablation depositions

The laser ablation deposition process can be divided into the following stages [6]: (i) heating, melting and fast evaporation of the target material by the laser beam; (ii) absorption of laser power in the evaporated material and the formation of plasma followed by an isothermal expansion; (iii) adiabatic expansion of the plasma and formation of a directed particle beam, and (iv) condensation of the material on the substrate and the nucleation and growth of the film. If the growth occurs on a non-heated substrate, the structure of the film is usually amorphous; crystalline films are

obtained either by post deposition-annealing, or in an in situ process on a heated substrate.

A typical laser ablation chamber is shown in Fig. 1 [1]. The laser beam is usually scanned by a computer-controlled xy-stage over the target surface at an angle of 30°–45°. A typical distance between the target and substrate is 30 to 40 mm. In some arrangements, the target can rotate and move up and down. The substrate can be heated to 800 °C. The base pressure of the chamber is 10^{-5} mbar; it is also possible to deposit films in low-pressure reactive gas ambients. Laser ablation deposition was carried out mostly by using a pulsed XeCl excimer laser (wavelength 308 nm, maximum pulse energy about 50 mJ, pulse duration about 20 ns and maximum repetition rate 50 Hz). In some experiments also a Q-switched Nd : YAG laser (wavelength 1064 nm, pulse energy 0.5–3 mJ, pulse duration 200 ns and repetition rate 60–1000 Hz) was used.

3. $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ superconducting films and $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{PrBa}_2\text{O}_{7-\delta}$ superlattices

$\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (YBCO) and $\text{PrBa}_2\text{O}_{7-\delta}$ (PrBCO) films, YBCO/PrBCO superlattices and $\text{Y}_{0.5}\text{Pr}_{0.5}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ films were deposited on heated (700–750 °C) SrTiO_3 single crystal substrates using an energy density of 2–3 J/cm² and an oxygen flow (0.25 mbar O₂) during the deposition [2]. For the deposition of superlattices the laser beam was scanned over two targets placed closely together. It was possible to control the thickness of each layer independently by

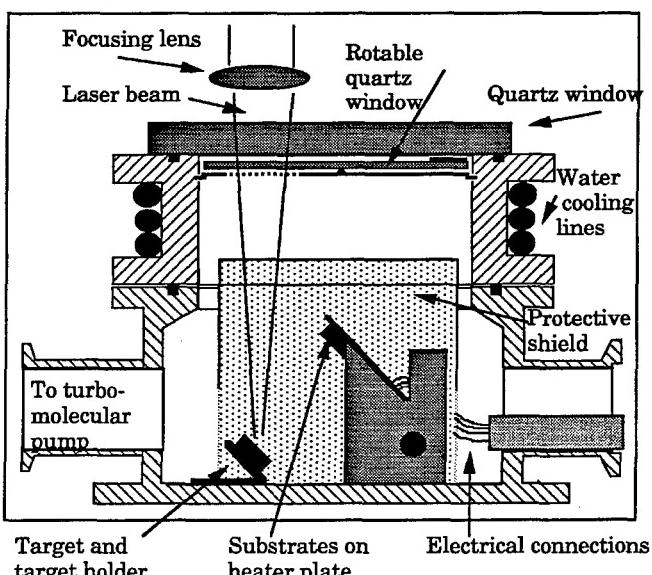


Fig. 1. A typical laser ablation chamber.

varying the number of pulses. Typical deposition rates were about 0.1 \AA/pulse ($2.5\text{--}5 \text{ \AA/s}$). The thin films and superlattices were epitaxial with the c-axis perpendicular to the SrTiO_3 (001) surface plane. The zero resistivity temperature, T_{c0} , of the epitaxial YBCO films was $87\text{--}90 \text{ K}$ and the relative resistance curve showed metallic behaviour above T_c and extrapolated well to zero (Fig. 2, curve a) [2]. The resistivity curve, $R(T)$, of the PrBCO film showed semiconductor-like behaviour.

For the 2/2 and 1/1 unit cell period superlattices (Fig. 2, curves b and c), the respective critical temperatures, 57 K and 31 K , are lower than for YBCO but considerably higher than 18 K for the alloy $\text{Y}_{0.5}\text{Pr}_{0.5}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ (Fig. 2, curve d). For longer period superlattices, the resistivity curves and critical temperatures approach those of YBCO films. All superlattice films showed metallic behaviour above T_c but the resistivity ratio, $R(300 \text{ K})/R(100 \text{ K})$, was considerably less than the value of 3 exhibited by the epitaxial film. Even the resistivity curve of the 1/1 unit cell superlattice had a positive slope contrary to that of the $\text{Y}_{0.5}\text{Pr}_{0.5}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$ material, and its T_{c0} was at least 12 K higher so that Y and Pr were not randomly distributed in the rare earth sites. The critical current density in a zero magnetic field at 77 K was in the range $1\text{--}2.5 \times 10^6 \text{ A/cm}^2$ for the YBCO films. For a 4/4 superlattice film the critical current density at 30 K was $0.2 \times 10^6 \text{ A/cm}^2$.

For hybridization of high T_c superconductors and semiconductors, silicon is the most important substrate. Deposition of good quality high T_c superconducting thin films directly on to silicon has proved to be impossible because the diffusion of silicon destroys the superconducting properties of the film. Therefore a buffer layer on a silicon substrate has to be used. In a study, laser ablation deposition of high T_c superconducting $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ films on silicon

using an yttrium-stabilised zirconia (YSZ) buffer layer was investigated. The substrates were silicon and silicon on sapphire. The films on the YSZ buffer layer were very promising: they were almost epitaxial, showed metallic type resistivity behaviour and had zero resistivity temperatures of about 85 K .

4. Lead zirconium titanate and PZT/YBCO heterostructures

Lead zirconium titanate (PZT) materials are of great interest for high dielectric-constant and non-volatile memory chips and for piezoelectric sensor and actuator applications. In LAD of PZT it is necessary to optimise the processing parameters for each material composition. For example, the laser fluence has a critical effect on the composition and microstructure of films. If too high a fluence is used, large particles impair the quality of the film. On the other hand, the target-substrate distance together with the laser fluence affects the growth rate of the films. The optimised value for the fluence at a deposition distance of 40 mm was $1\text{--}1.5 \text{ J cm}^{-2}$ for $\text{Pb}_{0.97}\text{Nd}_{0.02}(\text{Zr}_{0.55}\text{Ti}_{0.45})\text{O}_3$ targets [1]. To avoid ablation from a small area, a rotating target which moved up and down was used. After ablation on unheated sapphire substrates (with R-plane surfaces) the films ($100\text{--}800 \text{ nm}$ thick) were amorphous and had good surface morphology. After deposition the films were annealed in air with PZT-powders at 750°C for two hours. The films had a trigonal perovskite structure as the main crystal phase and the composition of the target was reliably reproduced.

In connection with PZT, high T_c superconducting oxides can be used as electrodes (in the normal state) or as an active superconducting material. LAD was used to produce layered structures containing PZT and YBCO-compounds [3]. The layers were deposited in situ on heated (PZT: $585\text{--}600^\circ\text{C}$, YBCO: $740\text{--}760^\circ\text{C}$) single crystal substrates (fluence $1\text{--}2 \text{ J/cm}^2$) in an oxygen flow. The highest crystal perfection of PZT was obtained on epitaxial YBCO deposited on a single crystal of strontium titanate.

5. Cu(In, Ga)Se₂

CuInSe_2 (CIS) and $\text{Cu}(\text{In}, \text{Ga})\text{Se}_2$ (CIGS) have been recognized as most important new semiconductor photovoltaic materials, with large-scale applications projected for thin film solar cells and radiation detectors. Gallium substitution of CIS provides a convenient means of adjusting the band gap of the material. As x goes from 1 to 0 the band gap of $\text{CuIn}_x\text{Ga}_{1-x}\text{Se}_2$ increases from approximately 1.0 eV to 1.68 eV . Applications for IR detectors would benefit from the small bandgap of CIS. For solar cell applications $\text{CuIn}_{0.75}\text{Ga}_{0.25}\text{Se}_2$ with a bandgap of 1.17 eV presents a useful compromise between the optimum bandgap of 1.5 eV and ease of film preparation.

Targets were prepared from polycrystalline and Bridgman-grown single crystal material [4]. The peak power density was 10^9 W/cm^2 . The films (typically 400 nm thick) were deposited at 350°C on fused silica and silicon (111) single crystal substrates. After deposition, the temperature was lowered to room temperature without any additional heat treatment. The EDAX results for the bulk analysis of the ablated films revealed that, in general, the selenium content was below that of the target material.

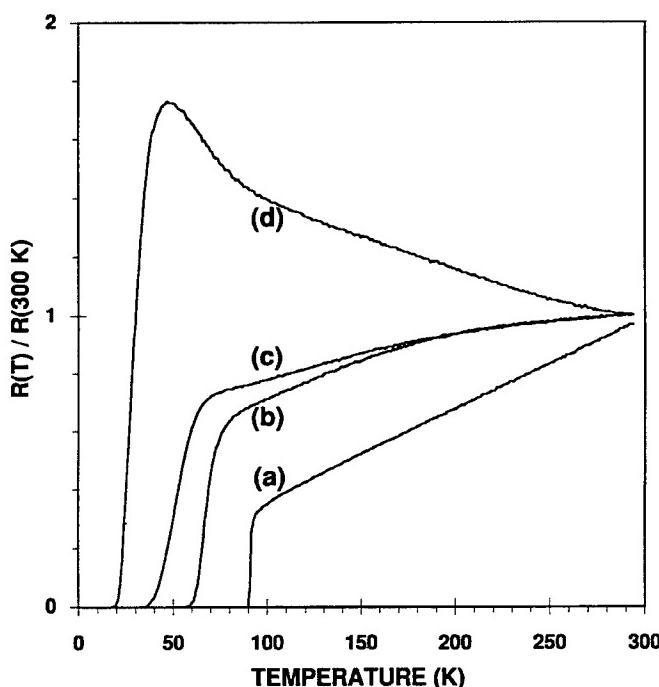


Fig. 2. Relative resistance, $R(T)/R(300 \text{ K})$, of films on SrTiO_3 : (a) $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ deposited by an in situ process and (b) $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}/\text{PrBa}_2\text{O}_{7-\delta}$ superlattice, period 2/2 unit cell (4.8 nm), thickness 97 nm (21 + 20 layers), (c) period 1/1 unit cell (2.4 nm), thickness 58 nm (25 + 24 layers), (d) alloy $\text{Y}_{0.5}\text{Pr}_{0.5}\text{Ba}_2\text{Cu}_3\text{O}_{7-\delta}$.

However, the indium/gallium ratios of the target materials were seen to be faithfully reproduced up to 20% gallium substitution. The films were all found to be n-type, as were the original targets.

The ability of LAD to prepare films whose bandgap may be varied simply by varying the composition of the starting material is of great interest in the field of fibre optic communications.

6. Diamond-like carbon thin films

Amorphous diamond-like carbon (DLC) thin films were deposited on fused silica and single crystal silicon substrates using a pyrolytic graphite target [5]. The thickness of the films was between 100 nm and 500 nm and the deposition time was 10–30 min.

The effect of substrate temperature and the intensity of the laser pulse was investigated. By varying the power density during the laser pulse between 10^9 and 10^{10} W/cm² and the deposition temperature between room temperature and 200 °C it was found to be possible to tailor the properties of the deposited films between the extremes of diamond-like and graphite-like carbon (Fig. 3). The lowest substrate temperature and the highest peak power density yielded the most diamond-like properties. The optical band gap of the films varied from zero (graphite-like films) to a maximum value of 1.35 eV (DLC) for films deposited at 200 °C and at

room temperature, respectively. As a function of the power density, the optical gap varied between 0.45 and 1.35 eV for films deposited at room temperature [Fig. 3(a)]. By introducing hydrogen into the chamber during deposition, the DLC band gap could be increased to 2.2 eV. The real part (n) of the refractive index was near 2.4 in the visible wavelength range for good quality unhydrogenated DLC films and near 1.4 for hydrogenated films.

The electrical resistivity of the unhydrogenated films could be varied over six orders of magnitude by adjusting the power density [Fig. 3(b)]. As a function of the deposition temperature between room temperature and 200 °C the variation was more than eight orders of magnitude [Fig. 3(c)].

The large variation of the optical and electrical properties of laser ablated DLC films as a function of deposition parameters opens up the possibility to utilise these films in electronics, either as active materials or as optical or protective coatings.

7. Conclusions

LAD is an attractive preparation method to realise electronic materials. It is specially suitable for multicomponent systems and for superlattices because of stoichiometry reproducibility and the ease with which layer thickness can be adjusted. Complicated material composition and even variation of the composition of the film in a scale of a unit cell can be realised. Good quality films of PZT and CIGS materials show that LAD is suitable to make films consisting of both more and less volatile compounds. LAD of DLC is an example of its application to a very high-melting point element. As a production method LAD is still at an early stage. There are problems with surface quality (because of sub micron-sized droplets on the film) and deposition of large areas with contrast thickness.

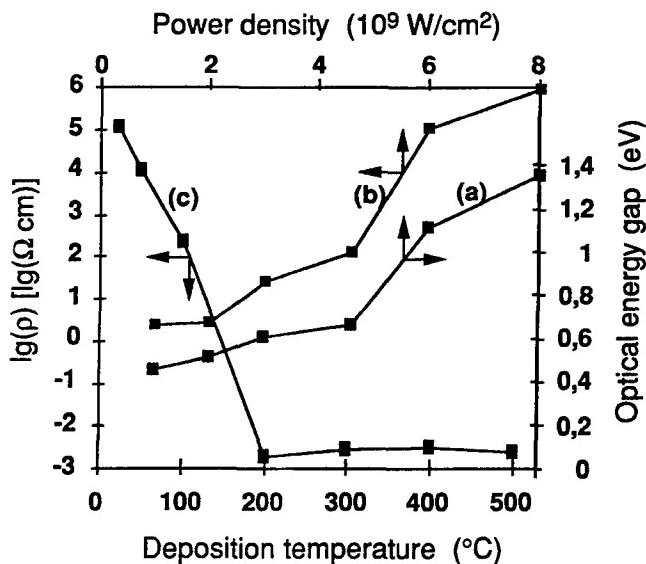


Fig. 3. Optical energy gap of a DLC film deposited at 20 °C as a function of laser power density (a) and logarithm of electrical resistivity as a function of the laser power density (b) and the deposition temperature (c).

References

1. Leppävuori, S. et al, Sensors and Actuators A41–42, 145 (1994).
2. Levoska, J. and Leppävuori, S., Active and Passive Electronic Components 15, 139 (1993).
3. Levoska, J., Leppävuori, S., Murtoniemi, and Lappalainen, J., Electroceramics IV Conf. in Aachen 1994.
4. Hill, A. E. et al, Mat. Res. Soc. Symp. Proc 285, 483 (1993).
5. Leppävuori, S. and Levoska, J., To be published in Conf. Proc. of IUMRS-ICAM-93.
6. Singh, R. and Narayan, J., Phy. Rev. B41, 8843 (1990).

Growth of Silicon Carbide on (100) Silicon Substrates by Molecular Beam Epitaxy

V. M. Airaksinen, J. Kaitila, H. Niemi¹, J. Lahtinen¹ and J. Saarilahti²

Electron Physics Laboratory, Helsinki University of Technology, Otakaari 7A, 02150 ESPOO, Finland.

Received May 5, 1994; accepted June 15, 1994

Abstract

A plasma activated gas source molecular beam epitaxy process has been developed in which the molecular beam is formed by activating a methane-hydrogen mixture in a plasma source. Amorphous carbon growth on (100) silicon substrates occurs when the substrate temperature exceeds 800 °C. The growth of cubic silicon carbide is observed above 880 °C. Epitaxial silicon carbide layers are characterised using X-ray photoemission spectroscopy, atomic force microscopy, ellipsometry and Rutherford backscattering.

1. Introduction

Cubic silicon carbide (SiC) and diamond are wide band gap semiconductors with many potential applications for high frequency, high power and high temperature devices [1]. The basic problem for the semiconductor applications of carbon is the absence of epitaxial methods for the growth of high quality single crystal diamond and Si-C alloy films on affordable substrates such as silicon. In recent years, however, considerable progress has been made with silicon carbide. The effort to develop diamond fabrication technology has been substantial, but it seems that the growth methods available at present are not suitable for the heteroepitaxial growth of diamond. Due to the large bonding and surface energy of diamond the growth does not proceed planarly but results in the formation of small crystallites [2].

The most common methods for the growth of diamond-like and diamond films are various plasma assisted chemical vapour deposition (CVD) techniques in which the precursor gases are activated by a plasma. The pressure during the process is typically quite high and the substrate is usually directly in contact with the plasma. The growth mechanism of diamond is still unclear because the number of possible chemical reactions is large and direct characterisation of the growth process is difficult.

Molecular beam epitaxy offers the potential for the successful heteroepitaxy of silicon carbide and diamond. In MBE the growth occurs in ultra high vacuum (UHV) from elements evaporated with thermal sources or from gaseous precursors. MBE has several advantages for the growth of carbon films: The growth at UHV; the substrate surface can be made atomically clean by heat cleaning methods; the substrate can be modified by using buffer layers; because the source(s) and the substrate are separated, no metastable

equilibrium is formed between the surface and the vapour phase and the chemical reactions on the surface are separate from the reactions occurring in the plasma; due to the good vacuum electron diffraction methods can be used for surface characterisation during the growth; and there are several independently controllable parameters available for the growth experiments (substrate temperature, substrate bias, gas flow rates, plasma input power).

Our aim is to develop an MBE method for the heteroepitaxial growth of diamond. SiC could provide a suitable substrate for the growth of epitaxial diamond on silicon. The growth of epitaxial SiC on silicon by MBE has previously been reported using electron beam evaporation [3, 4] and hydrocarbon gas sources [5, 6]. In this paper we report the use of a plasma source to activate the hydrocarbon gas for the growth of SiC on silicon.

2. Experimental

A modified VG Semicon V80M molecular beam epitaxy (MBE) system is used for the growth of carbon containing layers. A schematic diagram of the growth apparatus is shown in Fig. 1. Methane (CH_4) and hydrogen (H_2) are used as the source gases. Gas flow rates are controlled by mass

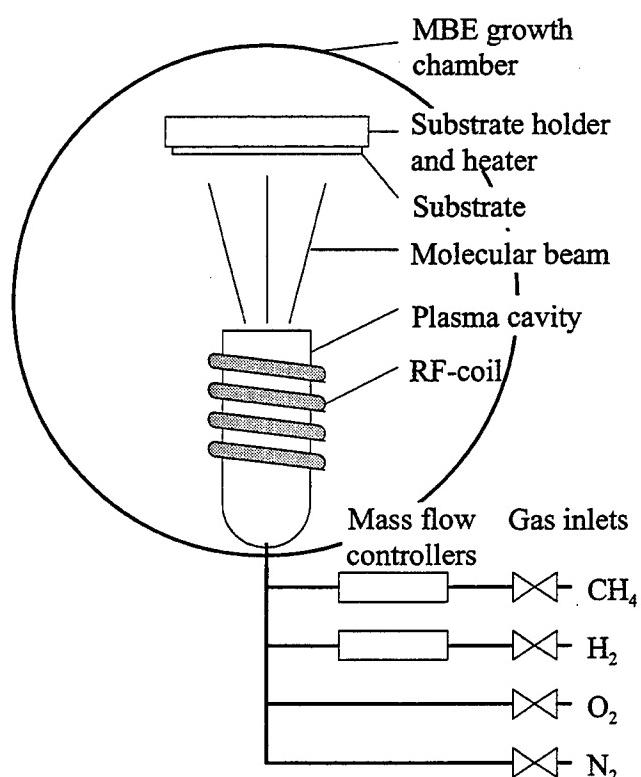


Fig. 1. Schematic diagram of the plasma assisted MBE growth apparatus.

¹ Present address: Laboratory of Physics, Helsinki University of Technology, Otakaari 1, 02150 ESPOO, Finland.

² Present address: Electronic Materials and Components, VTT Electronics, Technical Research Centre of Finland, Otakaari 7B, 02150 ESPOO, Finland.

flow controllers and the gases are mixed in a manifold before being injected into the plasma cell through a needle valve. A plasma activated gas cell (Oxford Applied Research MPD - 21) is used as the carbon source. The plasma source consists of a small cavity surrounded by the radio frequency (RF) coil. The end plate of the cavity contains a few small apertures (we have mostly used 9 holes, each 1 mm in diameter), through which the molecular beam is ejected into the growth chamber. The methane-hydrogen gas mixture is activated in the cavity using an RF-field. The substrate is located 360 mm from the plasma cell apertures, and it is heated by a resistive heater element. The substrate temperature is monitored by an optical pyrometer and a thermocouple located behind the substrate. The pumping system of the growth chamber consists of an oil diffusion pump and liquid nitrogen cooled cryopanels. The growth chamber is also equipped with a reflection high energy electron diffraction (RHEED) system for *in situ* monitoring of the growth. The other parts of our MBE-system consist of a growth chamber for III-V compounds and a preparation chamber. An Auger spectrometer for surface analysis and a quadrupole mass spectrometer for residual gas analysis are situated in the preparation chamber.

For the growth of SiC 75 mm diameter (100)-oriented silicon wafers are used as substrates. First the substrates are cleaned using a standard chemical cleaning process. The thin native oxide is removed *in situ* by heating the substrate above 1000 °C in the growth chamber. The oxide removal is observed using RHEED. Mass flow rates during the growth are typically 0.35 sccm and 3.5 sccm for methane and hydrogen, respectively. Substrate temperatures, as determined by the pyrometer, are varied between 700 °C and 970 °C. Typical power input to the RF-coil is between 100–165 W. Growth times have been between 20–45 min. After the growth sequence the plasma cell is cleaned with an oxygen (O₂) plasma. The background pressure in the MBE chamber is about 10⁻¹¹ mbar when the system is idling. During the growth the pressure rises to 10⁻⁴ mbar due to the hydrogen background buildup. The samples have been characterised using Rutherford backscattering spectroscopy (RBS), ellipsometry, reflection high energy electron diffraction (RHEED), X-ray photoelectron spectroscopy (XPS) and atomic force microscopy (AFM).

3. Results and Discussion

Growth parameters were varied to find the optimum conditions for the condensation of carbon on the substrate surface. The crucial parameter turned out to be the substrate temperature. Below 800 °C no growth occurred. Between 800 °C and 880 °C a layer of amorphous, hard carbon was grown and above 880 °C epitaxial SiC was formed. The RBS spectrum (Fig. 2) of a sample grown at 860 °C shows that there is a thin overgrowth of carbon on the Si substrate. The areal density of the carbon atoms is 2.64 × 10¹⁷ 1/cm². The amount of impurities (for example N or O) is very low. The concentration of atomic hydrogen was not measured.

The growth of epitaxial SiC occurs when the substrate temperature exceeds 880 °C. Epitaxy can be observed using RHEED. For layers grown below 880 °C the RHEED pattern is spotty and hazy. Above 880 °C, however, the

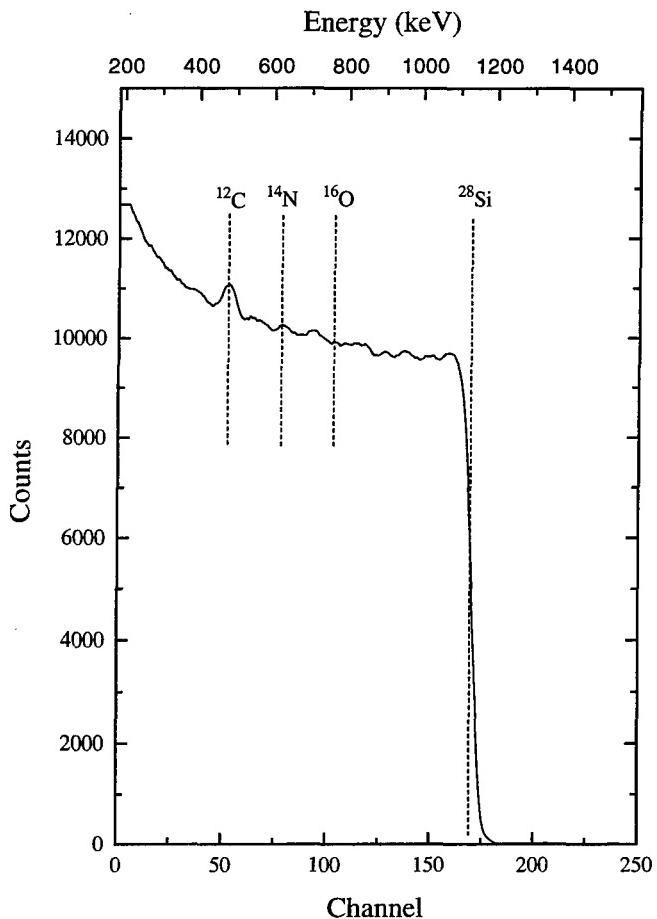


Fig. 2. RBS spectrum of a carbon layer grown at 860 °C measured using $^{12}\text{C}(\alpha,\alpha)^{12}\text{C}$ nuclear resonance.

RHEED pattern becomes streaky, bright and sharp indicating the formation of an epitaxial layer in a 3-dimensional growth mode. In Fig. 3 the RHEED pattern of a sample grown at 880 °C is shown in the <100> direction of the Si substrate. The analysis of the distances between the diffracting planes from the pattern indicates that the orientation between the SiC layer and the substrate is <110>SiC||<100>Si and <111>SiC||<110>Si. The lattice constant of the epitaxial layer obtained from the RHEED pattern agrees with that of SiC within the experimental error.

In order to confirm that the epitaxial layer is indeed SiC, the XPS spectrum was measured. As shown in Fig. 4, the spectrum shows clearly a peak at 283.20 eV which is characteristic of SiC [7]. The XPS spectrum also shows strong features at 284.8 eV and 286.1 eV, and weaker peaks at 287.7 and 289.2 eV, indicating the formation of C-C bonds. It is therefore likely that some graphite is also formed on the surface. For comparison the corresponding spectrum for a graphite control sample grown by vacuum evaporation is also shown. To gain a better understanding of the surface morphology of the SiC layer atomic force microscopy was used. The AFM showed that the size of the SiC crystallites ranges between 100–200 nm and they are epitaxially oriented along the <110> directions of the Si substrate. The thickness of the grown layers varies between 10–20 nm and the mean roughness is about 2.4 nm. The refractive index measured by single wavelength ellipsometry is 2.8, which is slightly lower than that of SiC probably due to the formation of voids between the crystallites.

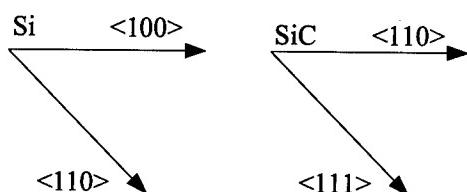


Fig. 3. The RHEED pattern of a SiC layer grown at 880°C taken in the $\langle 100 \rangle$ direction of the Si substrate.

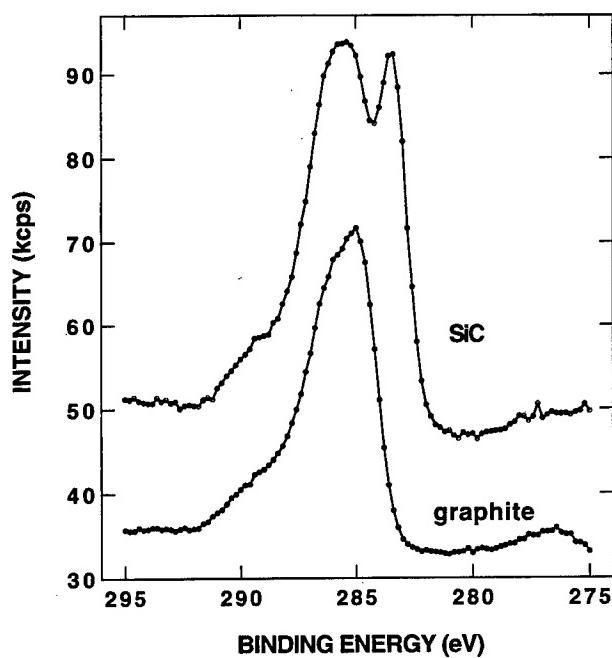


Fig. 4. XPS spectrum of a SiC layer grown at 880°C.

The composition of the molecular beam coming from the plasma cell is not known. However, it is easily shown that the activation of the methane/hydrogen mixture is essential to the growth. If the plasma is turned off, no growth is observed under any of the conditions used, in agreement with the results of Kusunoki *et al.*, who found that a CH₄ beam does not react with a Si substrate even at 1400 K [5]. Also, it seems likely that the carbon containing species in the molecular beam is not elemental carbon, because under identical conditions growth occurs on silicon but not on silicon oxide. The sticking coefficient of carbon cannot be determined accurately in our experiments. However, the plasma source is very nearly an ideal Knudsen cell [8] and by using the areal density of carbon obtained from RBS measurements, the lower limit of the sticking coefficient on silicon can be estimated to be about 0.65. Because some of the carbon is accumulated inside the plasma cavity, it seems that the true value of the sticking coefficient is nearly 1.

At the present time the main difficulty with our method is the carbon build-up inside the RF cavity. When the conducting layer of carbon on the cavity walls becomes sufficiently thick to shield the cavity from the RF field, the plasma is extinguished preventing the growth of thick layers.

In conclusion, we have achieved SiC growth on silicon by a plasma assisted MBE process. Our method shows promise for the growth of SiC and diamond on Si substrates because the growth parameters (the flux of carbon and other growth substituents, the plasma temperature, the substrate temperature and bias) can be independently controlled. This may provide valuable information about the growth mechanism of carbon.

References

1. Trew, R. J., Yan, J.-B. and Mock, P. M., Proc. IEEE **79**, 598 (1991).
2. Ravi, K. V. and Joshi, A., Appl. Phys. Lett. **58**, 246 (1991).
3. Kitabatake, M., Deguchi, M. and Hirao, T., J. Appl. Phys. **74**, 4438 (1993).
4. Zhou, G. L. *et al*, Cryst. J. Growth **134**, 167 (1993).
5. Kusunoki, I., Hiroi, M., Sato, T., Igari, Y. and Tomoda, S., Applied Surface Science **45**, 171 (1990).
6. Sugii, T., Aroyama, T. and Ito, T., J. Electrochem. Soc. **137**, 989 (1990).
7. Stoner, B. R., Ma, G.-H. M., Wolter, S. D. and Glass, J. T., Phys. Rev. B**45**, 11067 (1992).
8. See for instance Chapter 2 in "The Technology and Physics of Molecular Beam Epitaxy" (Edited by E. H. C. Parker) (Plenum Press, New York and London 1985).

Growth and Characterization of Compositionally Graded, Relaxed $\text{Si}_{1-x}\text{Ge}_x$

Arne Nylandsted Larsen, John Lundsgaard Hansen, Ralf Schou Jensen, Sergey Y. Shiryaev, and Peter Riis Østergaard

Institute of Physics and Astronomy, Aarhus University, DK-8000 Aarhus C, Denmark

Joachim Hartung and Gordon Davies

Physics Department, King's College London, Strand, London WC2R 2LS, United Kingdom

and

Flemming Jensen and Jon Wulff Petersen

Mikroelektronik Centret, The Technical University of Denmark, DK-2800 Lyngby, Denmark

Received April 28, 1994; accepted and revised form June 27, 1994

Abstract

Compositionally graded, relaxed, n-type, $\text{Si}_{1-x}\text{Ge}_x$ alloy layers have been grown on (100) Si substrates; the main emphasis has been put on compositions with $x = 0.25$. It is found that for substrate growth-temperatures higher than $\sim 750^\circ\text{C}$ and a grading rate of 10% Ge/ μm relaxed $\text{Si}_{0.75}\text{Ge}_{0.25}$ epitaxial layers of high structural, optical, and electrical quality can be grown. The layers are characterized by channeling parameters close to expected bulk values, a threading dislocation density of $\sim 5 \times 10^5 \text{ cm}^{-2}$, and strong near-band gap luminescence. Electrical measurements have revealed Hall mobilities similar to published bulk values and concentrations of electrically active deep levels $\leq 2 \times 10^{11} \text{ cm}^{-3}$. The surface morphology is, however, strongly influenced by the grading procedure which produces a high degree of cross-hatching.

1. Introduction

SiGe alloy layers have lattice constants which are up to 4% larger than the one for Si depending on the composition. Sufficiently thin $\text{Si}_{1-x}\text{Ge}_x$ layers can grow pseudomorphically, and thus strained, onto a silicon substrate. If the layer thickness exceeds a composition-dependent critical thickness, however, it relaxes structurally by the introduction of dislocations having misfit segments parallel to the interface and threading segments through the epitaxial layer [1]. Threading segment densities of $\sim 10^{11} \text{ cm}^{-2}$ are formed for $\text{Si}_{1-x}\text{Ge}_x$ with $x = 0.25$ [2], rendering this material useless for most applications. For a $\text{Si}_{0.75}\text{Ge}_{0.25}$ layer the critical thickness is $\sim 500 \text{ \AA}$.

Recently, however, it has been shown that relaxed, epitaxial $\text{Si}_{1-x}\text{Ge}_x$ layers of any composition can be grown on (100)-Si substrates with a very low density of threading dislocations by molecular beam epitaxy (MBE) or chemical vapour deposition (CVD) [3, 4]; threading dislocation densities between 1×10^4 and $1 \times 10^6 \text{ cm}^{-2}$ in the top layer have been reported. This has been achieved by growing the top layer of the required composition on a compositionally graded buffer layer in which the Ge content has been increased gradually. The reduction of the density of threading dislocations from $\sim 10^{11}$ to $\sim 10^4 \text{ cm}^{-2}$ for a $\text{Si}_{0.75}\text{Ge}_{0.25}$ layer by the compositional grading procedure is explained as a consequence of a self-adjustment of the dislocation multiplication sources in the $\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ hetero-

epitaxial system, resulting in an increased probability for the threading dislocations to annihilate [5].

Due to their potential use for the fabrication of novel devices as well as for studies of bulk properties of unstrained $\text{Si}_{1-x}\text{Ge}_x$ alloys these epitaxial layers have received a great deal of attention. Lattice-matched III-V layers on Si chips have already been grown with these relaxed buffer layers as templates [4], and very high electron mobility, two-dimensional electron gas has been achieved in Si by growing on a relaxed, graded SiGe buffer on Si [6, 7].

In the present article we report on the characterization of epitaxial, compositionally graded $\text{Si}_{1-x}\text{Ge}_x$ grown by molecular beam epitaxy (MBE). The main emphasis will be put on compositions with $x = 0.25$, which are either unintentionally or n-type doped. A number of experimental techniques have been utilised revealing structural, optical, and electrical information.

2. Experiment

The $\text{Si}_{1-x}\text{Ge}_x$ epilayers were grown by MBE on 100 mm (100) Si substrates in a VG Semicon V80 system. The base and growth pressures were 5×10^{-11} and about 5×10^{-10} Torr, respectively. The total (Si plus Ge) growth rate of 5 \AA/s was maintained constant throughout all the growth procedures whereas the substrate temperature was varied between 540 and 850°C . A silicon buffer layer of thickness $1 \mu\text{m}$ was first grown followed by the growth of the graded buffer layer with a starting Ge content of about 0.2%, and with a Ge grading rate of 10% Ge/ μm . The thickness of the top uniform layer was 3–4 μm and the top layer was either unintentionally doped in which case it was n-type with a carrier concentration $\leq 2 \times 10^{14} \text{ cm}^{-3}$ or n-type doped with Sb to a concentration of $\sim 1 \times 10^{16} \text{ cm}^{-3}$. The compositions of the layers were determined by Rutherford back-scattering spectrometry (RBS) using 2-MeV He^+ -ions. The crystalline quality of the SiGe layers was studied by RBS/channeling, planar and cross sectional transmission electron microscopy (TEM) using a Philips CM 20 microscope operating at 200 kV, and Schimmel etch-pit test. The optical properties were investigated by photoluminescence (PL)

using an Ar^+ ion laser with an excitation power between 5–50 mW/mm²; the samples were mounted stress-free and were immersed in liquid helium, $T = 4.2$ K; the PL spectra were recorded with a Nicolet 60SX Fourier transform spectrometer fitted with a North Coast Ge diode detector. Hall mobilities were measured by the van der Pauw method in the temperature range from 80 to 300 K on specially grown samples in which the n-type top layer was separated from the graded buffer layer by a pn-junction; Au(99%)/Sb(1%) alloy contacts were evaporated in the corners of the square samples. Deep level transient spectroscopy (DLTS) with a SemiTrap DLS82 instrument was used to measure the concentration of electrically active deep centers in the band gap. Diodes for DLTS were fabricated either by mesa-etching samples which had been terminated by the growth of a 0.5 μm B-doped $Si_{0.75}Ge_{0.25}$ p⁺-layer or by Pd-evaporation through a diode forming mask followed by a sintering at about 300 °C for 20 s; the diodes were evaluated by measurements of current-voltage and capacitance-voltage characteristics prior to the DLTS measurements. Finally, the surface morphology of the layers was studied by atomic force microscopy (AFM) in air with a Rasterscope 3000 instrument.

3. Results and discussion

Figure 1 shows an example of a cross-sectional two-beam TEM image of a $Si_{0.75}Ge_{0.25}$ layer grown at 550 °C. The 2.5 μm thick compositionally graded layer with the high density of misfit dislocations can be easily distinguished from the 3 μm thick “dislocation free” uniform top layer; the characterization of the layer as being dislocation free should be taken with caution as TEM is not sensitive to dislocation densities less than $\sim 10^6$ cm⁻². To get a better estimate of the dislocation density we used the Schimmel etch-pit test which revealed dislocation densities of $\sim 5 \times 10^5$ cm⁻². It should be noted that the etch-pit test is also subjected to a large uncertain due to the rough surface morphology of the layers (see later). In this way the samples which are discussed in the following have been estimated to contain $\sim 5 \times 10^5$ cm⁻² dislocations.

Channeling angular scans of unintentionally doped $Si_{0.75}Ge_{0.25}$ alloy layers grown at different temperatures are

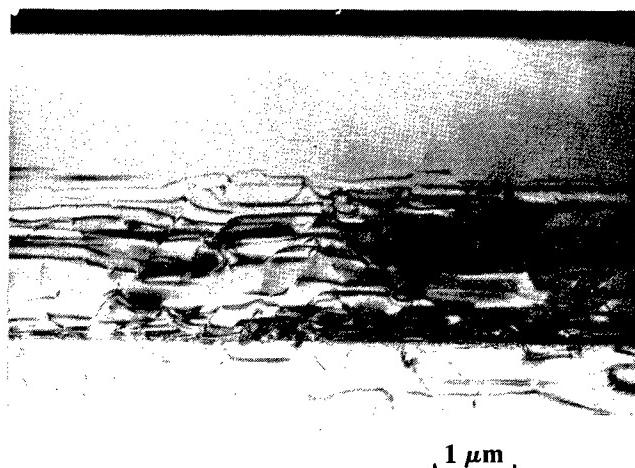


Fig. 1. Cross-sectional TEM micrograph of a $Si_{0.75}Ge_{0.25}$ structure grown at 550 °C. The compositional graded layer is 2.5 μm thick, and the top layer 3 μm .

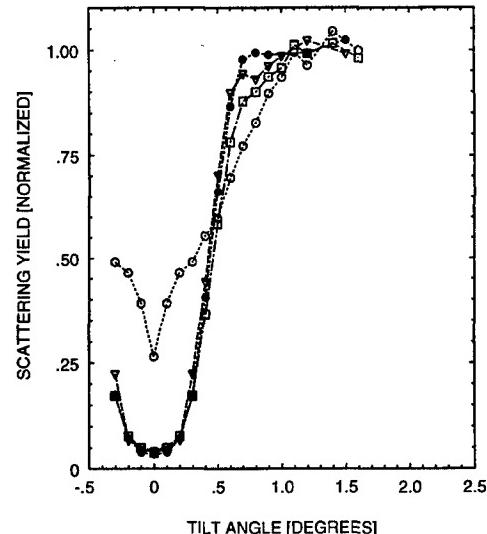


Fig. 2. Channeling angular scans around the <100>-direction measured with 2-MeV He⁺ particles. Virgin Si (●), $Si_{0.75}Ge_{0.25}$ grown at 540 °C (○), 750 °C (▽), and 850 °C (□).

displayed in Fig. 2 together with an angular scan of a virgin Si crystal; the results are collected in Table I. The scans were measured in a narrow depth-window of 500 Å at a depth of ~ 1000 Å. Whereas the scans of the samples grown at 750 and 850 °C resemble the one for virgin Si, the scan of the 550 °C-sample is sharpened at small tilt angles and broadened at larger tilt angles. The broad feature might appear as a consequence of a mosaic structure in the top layer caused by the network of misfit dislocation in the graded layer as previously reported for similar samples grown at 500–560 °C [2]. If a mosaic structure is assumed to cause the broadening, the mosaic angular spread can be estimated from the measured critical angles of Table I, assuming a gaussian distribution of the angular spread; this results in an angular spread of $\sim 0.5^\circ$. However, the form of the scan cannot be explained alone by a mosaic structure; a more complicated, and at present unknown, defect-structure results in the sharpening of the scan. It should be noted that TEM of samples grown at 550 °C did not reveal any defects. The expected value of $\psi_{1/2}$ for virgin Si is 0.45° [8] in agreement with the measured value, and for Ge 0.57° [8]. An interpolation for $Si_{0.75}Ge_{0.25}$ gives 0.47° which agrees well with the measured value of 0.46° of the 850 °C-sample; the value of the 750 °C-sample is, however, significantly smaller. This narrowing could be due to a composition modulation in depth in which the Ge concentration varies between 0.20 and 0.30 with a period of about 50 Å. Such a composition modulation has been observed for growth temperatures below 750 °C and is due to the geometrical arrangement of the e-guns in the MBE growth chamber. A narrowing could

Table I. Results from 2-Mev He⁺ channeling measurements, χ_{min} is measured along the <100> direction and $\psi_{1/2}$ is the critical angle for channeling in the <100>-direction

Sample	χ_{min}	$\psi_{1/2}$
Virgin Si	$3.5 \pm 0.5\%$	$0.44 \pm 0.01^\circ$
$Si_{0.75}Ge_{0.25}$, $T_g = 540$ °C	25.8%	$\sim 0.6^\circ$
$Si_{0.75}Ge_{0.25}$, $T_g = 750$ °C	$4.2 \pm 0.5\%$	$0.42 \pm 0.01^\circ$
$Si_{0.75}Ge_{0.25}$, $T_g = 850$ °C	$3.8 \pm 0.5\%$	$0.46 \pm 0.01^\circ$

also be the result of a weak mosaic structure giving rise to some minor dechanneling without any broadening of the angular scan; this would be in agreement with the slightly larger χ_{\min} value observed for the 750 °C sample compared to the 850 °C sample. Nevertheless, it can be concluded that for a growth temperature of 850 °C there is no indication of a mosaic structure with an angular spread of more than ~0.1° which is the detection limit of the present experiment.

The influence of the growth temperature on the photoluminescence is shown in Fig. 3. The PL measurements reported here were done on samples from the same wafers as were used for the channeling measurements. The PL spectra of the $\text{Si}_{0.75}\text{Ge}_{0.25}$ grown at temperatures higher than 750 °C are dominated by two peaks which are described as a no-phonon transition (NP) and its transverse optical (TO) Si-Si phonon replica according to their separation in photon energy. These peaks labeled X^{NP} and X^{TO} in Fig. 3(b) and (c) have been interpreted as excitons bound to unintentionally incorporated shallow impurities in SiGe alloys [9, 10]. In addition to this near-band gap luminescence the D-bands, D1–D4, which are related to impurity-decorated dislocations [11, 12], show up in the PL-spectra of the SiGe layers grown at temperatures higher than 750 °C, Fig. 3(b) and (c). The photon energies of the D-bands are reported to decrease in $\text{Si}_{1-x}\text{Ge}_x$ with increasing x [13, 14]. Due to different shift rates with the Ge-composition the D2-and D3-bands overlap at $x \geq 20\%$ [12]; as a consequence the D2 band in Fig. 3(b) is broadened as the less intense D3-band overlaps. The spectral positions of the D-bands recorded in spectra (c) and (b) give evidence that the underlying dislocations are located in the $\text{Si}_{0.75}\text{Ge}_{0.25}$ -layer. Compared with the spectra of the 750 and 850 °C samples the luminescence signal of the 540 °C sample shown in Fig. 3(a) is very weak. It consists of two broad overlapping bands centered at 1.0 eV and 0.79 eV. The latter one

continues beyond the low frequency cut-off of the Ge-detector. The observation of a similar broad band in fully strained epitaxial SiGe alloys grown at 600 °C has been reported in the literature and has been related to small interstitial-type platelets [15]. This would be consistent with the lack of any observable defects by TEM in samples grown at 550 °C. The absence of well resolved near-band gap luminescence in Fig. 3(a) indicates a rather low quality of the sample grown by MBE at 540 °C in agreement with the channeling results.

An example of a DLTS spectrum of an unintentionally doped $\text{Si}_{0.75}\text{Ge}_{0.25}$ layer grown at 750 °C is given in Fig. 4. The sample has been cut from the same wafer as was used for the channeling (Fig. 2) and PL (Fig. 3) measurements. The n-type doping was derived by capacitance-voltage (CV)-measurements to be $2 \times 10^{14} \text{ cm}^{-3}$. Due to the small thickness of the top $\text{Si}_{0.75}\text{Ge}_{0.25}$ layer of 4 μm a reverse voltage of only -1.5 V could be used before the depletion layer extended into the graded buffer layer. The broad peak centered around 180 K is probably due to dislocations decorated with impurities. The very low observed deep level concentration $\leq 2 \times 10^{11} \text{ cm}^{-3}$ is in agreement with the low concentration of impurity-decorated dislocations observed by PL which is sensitive to impurities decorating dislocations at concentrations higher than $\sim 1 \times 10^{11} \text{ cm}^{-3}$ [16]. An increase of the growth temperature to 850 °C did not result in a significant change of the deep level concentration. We have, however, observed that an Sb doping from an effusion cell results in an increased concentration of deep levels; for an Sb doping level of $1-5 \times 10^{16} \text{ Sb}/\text{cm}^3$ a deep level concentration of $\sim 1 \times 10^{13} \text{ cm}^{-3}$ is found; we are presently investigating possible reasons for this.

The Hall mobility of a $\text{Si}_{0.75}\text{Ge}_{0.25}$ sample grown at 750 °C and doped to $1 \times 10^{16} \text{ Sb}/\text{cm}^3$ is shown in Fig. 5 as a function of temperature. The room temperature value of $300 \text{ cm}^2/\text{Vs}$ is in good agreement with a literature value of $350 \text{ cm}^2/\text{Vs}$ [17]. An increase of the mobility of about 20% at low temperature was observed upon a furnace annealing at 900 °C for 15 min. This is in agreement with PL measurements of highly doped layers revealing a broad luminescence peak centered around 0.9 eV which disappears

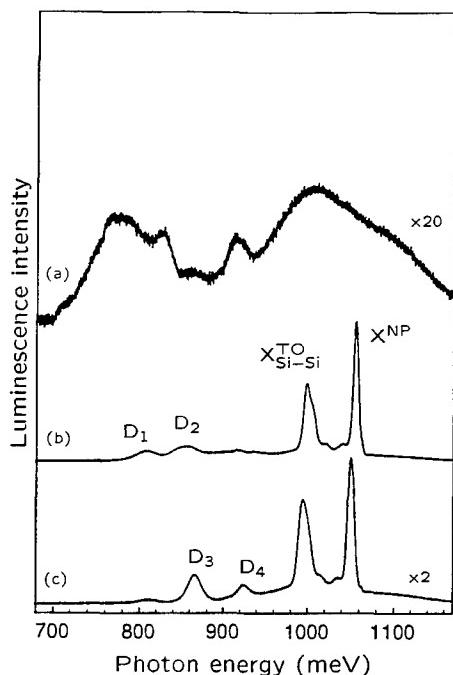


Fig. 3. Photoluminescence spectra of three unintentionally doped $\text{Si}_{0.75}\text{Ge}_{0.25}$ samples grown at (a) 540 °C, (b) 750 °C, and (c) 850 °C. The spectra were excited with the 488 nm line of an Ar^+ ion laser and an excitation power of 20 mW/mm².

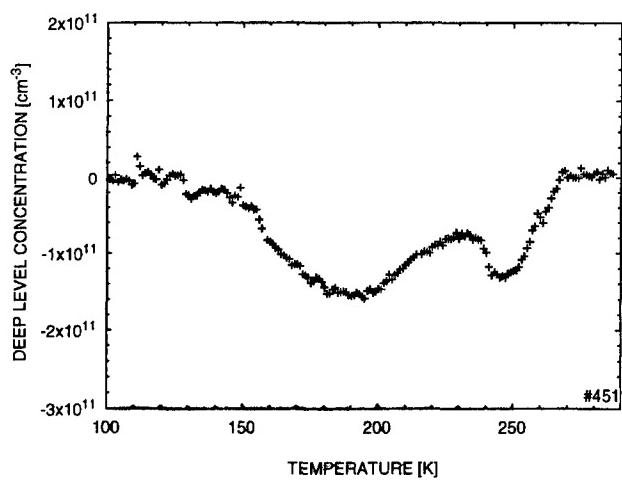


Fig. 4. DLTS spectrum of an unintentionally doped $\text{Si}_{0.75}\text{Ge}_{0.25}$ Schottky diode grown at 750 °C. The n-type doping concentration was $2 \times 10^{14} \text{ cm}^{-3}$. The spectrum was measured with a repetition frequency of 250 Hz and a pulse width of 50 μs.

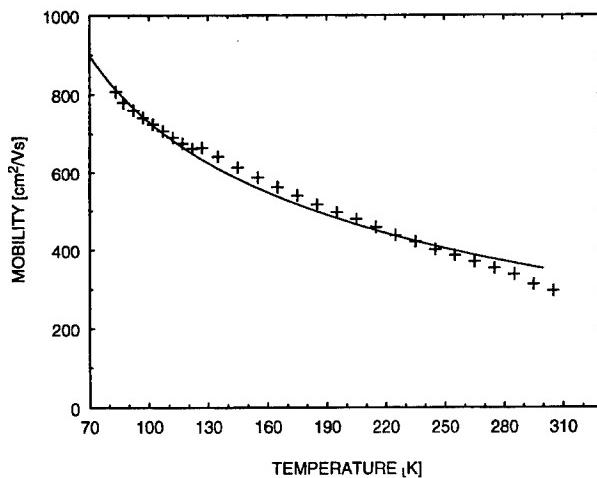


Fig. 5. Hall mobility as a function of temperature for a $Si_{0.75}Ge_{0.25}$ sample grown at 750 °C doped to $1 \times 10^{16} Sb/cm^3$. The full curve is a preliminary, theoretical estimate based on Brooks theory.

following a furnace annealing at 900 °C for 30 min. The temperature dependence is well described by the Brooks theory which includes the effect of alloy scattering [18].

Three-dimensional AFM images of typical surface morphologies of $Si_{1-x}Ge_x$ films grown at 550 and 750 °C are shown in Fig. 6. In both images clear cross-hatch patterns in the <110>-directions are apparent. The surface morphology is slightly smoother in the 750 °C-grown samples, and the surface roughness as measured by the root-mean-square roughness is also smallest for the 750 °C grown sample,

7.6 nm compared to 13.6 nm for the 550 °C grown sample. Thus, an increase of the growth temperature from 550 to 750 results in halving of the surface roughness. The appearance of these cross-hatch patterns on the surface of compositionally graded, epitaxial layers is presently being discussed as a result of the misfit-dislocation pattern in the compositionally graded layer separated from the surface by 4 μm of defect-free material [19, 20]: the surface morphology is expected to arise from a combination of both dislocation-induced mechanical shears of the surface and from growth phenomena controlled by local strain fields associated with misfit dislocations.

4. Conclusion

Compositionally graded, relaxed $Si_{0.75}Ge_{0.25}$ grown by MBE has been characterized by a combination of complementary analytical techniques. It is demonstrated that relaxed $Si_{0.75}Ge_{0.25}$ of high structural, electrical, and optical quality can be grown at temperatures higher than about 750 °C. The surface of these layers, however, is dominated by a cross-hatch pattern with a roughness of ~8 nm. For future use of this material for devices the issue of how to avoid this cross-hatch pattern has to be addressed.

Acknowledgements

This work was supported by the Danish Natural Research Council, the Danish Technical-Science Research Council, and the Danish National Research Foundation through the Aarhus Center of Advanced Physics (ACAP). Measurements in Kings' College London were supported by the Science and Engineering Research Committee.

References

- Fitzgerald, E. A., Mater. Sci. Rep. **7**, 87 (1991).
- Mooney, P. M., LeGoues, F. K., Chu, J. O. and Nelson, S. F., Appl. Phys. Lett. **62**, 3464 (1993).
- LeGoues, F. K., Meyerson, B. S., Morar, J. F. and Kirchner, P. D., J. Appl. Phys. **71**, 4230 (1992).
- Fitzgerald, E. A. et al., J. Vac. Sci. Technol. **B10**, 1807 (1992).
- Shiryayev, S. Yu., Phil. Mag. Lett. **68**, 195 (1993).
- Mii, Y. J. et al., Appl. Phys. Lett. **59**, 1611 (1991).
- Schäffler, F., Tobben, D., Herzog, H.-J., Abstreiter, G. and Hollander B., Semicond. Sci. Technol. **7**, 260 (1992).
- Feldman, L. C., Mayer, J. W. and Picraux, S. T., in: "Materials Analysis by Ion Channeling" (Academic Press, New York 1982), p. 43.
- Weber J., and Alonso, M. I., Phys. Rev. **40**, 5683 (1989).
- Higgs, V., Lightowers, E. C., Fitzgerald, E. A., Xie, Y. H. and Silverman, P. J., J. Appl. Phys. **73**, 1952 (1993).
- Sauer, J. et al., Appl. Phys. A**36**, 1 (1985).
- Higgs, V., Lightowers, E. C., Norman, C. E. and Kightley, P., Mat. Sci.-Forum **83-87**, 1309 (1992).
- Weber, J. and Alonso, M. I., in: "Defect Control in Semiconductors" (Edited by K. Sumino) (Elsevier Science Publisher B. V., North Holland 1990), p. 1453.
- Davies, G., et al., Mat. Res. Symp. Proc. **220**, 321 (1991).
- Noël, J. P., Powell, N. L., Houghton, D. C., Wang, A. and Perovic, D. D., Appl. Phys. Lett. **61**, 690 (1992).
- Lightowers E. C. and Higgs, V., Phys. Stat. Sol (a) **138**, 665 (1993).
- Busch, G. and Vogt, O., Helv. Phys. Acta **33**, 437 (1960).
- Azhdarov, G. Kh., Agaev, N. A. and Kyazimzade, R. A., Sol. Stat. Comm. **84**, 445 (1992).
- Hsu, J. W. P., Fitzgerald, E. A., Xie, Y. H., Silverman, P. J. and Cardillo, M. J., Appl. Phys. Lett. **61**, 1293 (1993).
- Shiryayev, S. Yu., Jensen, F. and Wulff Petersen, J., Appl. Phys. Lett. **64**, 3305 (1994).

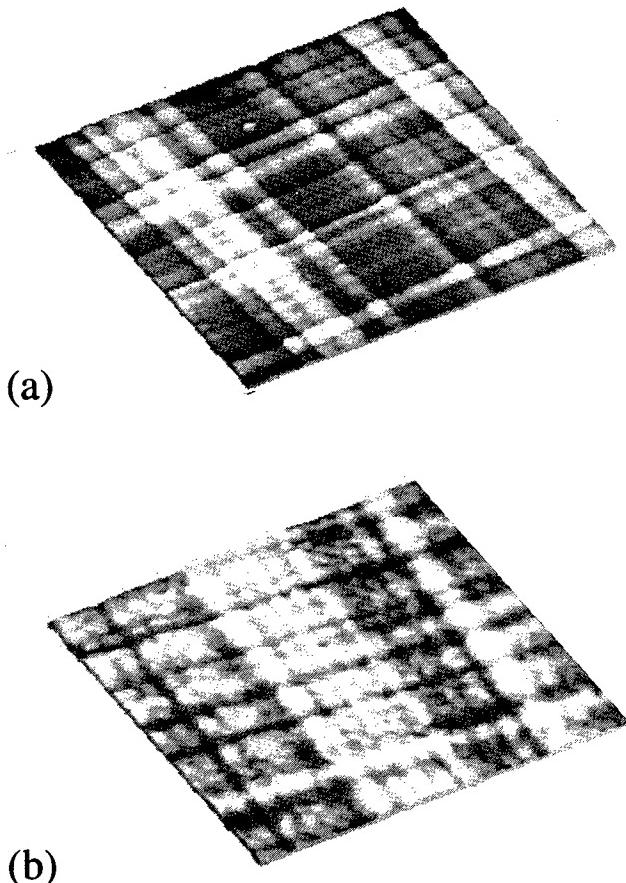


Fig. 6. AFM images of two $Si_{0.75}Ge_{0.25}$ samples grown at (a) 550 °C and (b) 750 °C. The scanned area is $17.5 \times 17.5 \mu m^2$ in both (a) and (b) and the z-scale (black to white) 81.1 nm in (a) and 53.8 nm in (b).

RBS Channeling Spectroscopy of Ge Implanted Epitaxial $\text{Si}_{1-x}\text{Ge}_x$ Layers

J. Saarilahti, Z. Xia, H. Ronkainen, P. Kuivalainen and I. Suni

VTT Electronics, Electronic Materials and Components, P.O. Box 1101, Fin-02044 VTT, Finland

Received April 28, 1994; accepted June 23, 1994

Abstract

$\text{Si}_{1-x}\text{Ge}_x$ layers were formed through high-dose germanium ion implantation into (100)Si substrates. Two alternative implantation techniques along with that of the single-energy Ge^+ implantation were separately adopted: the double-energy Si^+ and Ge^+ method, and the double-energy Ge^+ and Ge^{++} method. The purpose of the both double-energy methods was to form deeper amorphous layers by using relatively low-dose Si^+ or Ge^{++} ion bombardment while the SiGe alloy layers were created by high dose Ge^+ ion implantations. Furthermore, all the amorphized samples were epitaxially regrown by conventional furnace annealing at temperature of 525 to 600 °C. RBS channeling spectroscopy was used for optimizing these implantation processes. Measurements confirm that the double-energy Ge^+ and Ge^{++} method is optimum because of generating fewer residual defects. Additionally, the preliminary result on the regrowth properties of the double-energy Ge^+ and Ge^{++} implanted SiGe layer is also presented.

1. Introduction

Recently, the $\text{Si}_{1-x}\text{Ge}_x/\text{Si}$ heterojunction materials have been a focus of research because they greatly enhance the speed and flexibility of silicon and are compatible with existing silicon technology rather than replacing it. Using molecular beam epitaxy (MBE) and chemical vapor deposition (CVD), the cut-off frequency of the $\text{Si}_{1-x}\text{Ge}_x$ base heterojunction bipolar transistors (HBT) has jumped from 75 GHz in 1989 to the present 110 GHz [1, 2]. Typically, MBE and CVD have compositionally abrupt interfaces, but recent HBT studies suggest that compositionally graded structures may have advantages for some applications [1, 3]. Within this consideration, an approach has been proposed to use high dose germanium ion implantation into a single crystal silicon [4–9]. The resulting germanium distribution in the silicon near-surface region is Gaussian in shape, and the tail of the germanium profile extends beyond the base region. Numerical simulations indicate that if the position of the germanium profile is chosen properly, the high current gain is achievable also in the case of the germanium implantation [10]. Furthermore, this technology offers the opportunity to use $\text{Si}_{1-x}\text{Ge}_x$ in a selective area.

In this work we will represent RBS channeling results on solid phase epitaxy (SPE) of $\text{Si}_{1-x}\text{Ge}_x$ layers that were amorphized by either the double-energy Si^+ and Ge^+ implantation or the double-energy Ge^+ and Ge^{++} implantation as well as the usual single-energy Ge^+ implantation. In addition, the preliminary results will be also presented to reveal the regrowth properties of the double-energy Ge^+ and Ge^{++} implanted sample.

2. Experimental

2.1. Sample preparation

Cz-grown n-type 1–6 Ω cm Si wafers of ⟨100⟩ orientation were used throughout the experiments. The single-energy

Ge implantation was performed with 180 keV Ge^+ ions at a dose of $3.8 \times 10^{16} \text{ cm}^{-2}$. In regard to the double-energy Si^+ and Ge^+ implantation, the wafer was implanted first by 180 keV Si^+ ions at a dose of $1.0 \times 10^{16} \text{ cm}^{-2}$, and then by 180 keV Ge^+ ions at a dose of $4.7 \times 10^{16} \text{ cm}^{-2}$. Similarly, the double-energy Ge^+ and Ge^{++} implantation adopted a similar Ge^+ implantation as the former except that the Ge^{++} implantation of 360 keV is at a dose of $4.0 \times 10^{15} \text{ cm}^{-2}$. The main purpose of the both double-energy implantation methods employing Si^+ and Ge^{++} implantations was to form deeper amorphous layers while the SiGe alloy layers were created by 180 keV Ge^+ ion implantations. For suppressing the dynamic annealing effect which has been known to be detrimental for recrystallization of implanted layers, the amorphization implantation beam power densities used were less than 0.09 W/cm² for the high dose Ge^+ implantation and less than 0.46 W for the relatively low dose Si^+ or Ge^{++} implantation [11]. The corresponding wafer temperatures during the implantation were less than 200 °C and 380 °C, respectively. Solid phase epitaxial regrowth employing conventional furnace annealing processes was used to crystallize the amorphous layers.

2.2. Measurement techniques

The implanted samples were characterized by Rutherford Backscattering Spectroscopy (RBS) channeling measurements, which were carried out using a standard setup of a NEC 1.7 MV Tandem Pelletron Model 5SDH-2 accelerator. The 2000 keV ${}^4\text{He}^+$ analysing ion beam had angular beam divergence of less than 0.05 degrees, beam spot size of 1 mm² and a beam current of 25 nA on the sample at normal incidence. The backscattered ions were detected at a 170 degrees angle from a 100 mm distance of the measured sample at normal incidence by a standard 50 mm² silicon surface barrier detector with an energy resolution of 13 keV. With a separate detector connected to a single channel analyser, the ion doses were measured by counting the backscattering yield from a tantalum coated beryllium beam chopper rotating in the analyzing beam. The conversion from the beam chopper counts to ion beam dose was done to standard reference samples with known composition. The aligned measurements were performed along the ⟨100⟩ direction. The alignment was adjusted by a high precision two axis computer controlled stepper motor driving channeling goniometer. Counting rates and channeling ratio were determined directly from the spectra measured by the multichannel analyser (MCA).

The RBS spectra were solved numerically by the simulation program GISA-3 developed at the Technical Research Centre of Finland [13]. RBS spectra for both random and

channeled orientations of the measured samples were solved numerically. For the channeled spectrum, the amount of dechanneling is evaluated from the multiple scattering model [15]. Three parameters were needed for the crystal characterization and four parameters for the modelling of the damage distribution. The crystal quality was characterized by the minimum yield, dechanneling factor and the interatomic spacing between the crystal layers on the channeling direction. In order to have an excellent agreement between the Monte Carlo simulations performed by the TRIM-91 program [16, 17] and the GISA-3 spectrum analysis the stopping power formulation and parameters for the GISA-3 were adopted from the TRIM-91 program code. In addition, the stopping power values used for the GISA calculations for the incoming path of the projectile ions were modified by a multiplier (approximately 0.75) in order to take the channeling effect in account. For the channeling spectra of the damaged crystal samples the defect distribution was calculated from two jointed Gaussian distribution functions.

3. Results and discussion

With a single-energy Ge^+ implantation, the Ge distribution in a Ge implanted $Si_{1-x}Ge_x$ alloy layer is defined mainly by the projected ion energy and dose, and the generated amorphous layer thickness is also determined by these two parameters if dynamic annealing during implantation is suppressed. The regrowth of the amorphous layer is known to occur by the motion of the amorphous/crystalline (a/c) interface (not by the nucleation of new crystals within the amorphous layer) and usually defects classified as end-of-range (EOR) damage beyond the a/c interface are associated with the SPE growth of the amorphous layer [12]. If the Ge profile maximum is near the a/c interface and a p-n junction is located around the Ge profile maximum in order to satisfy the SiGe/Si bandgap engineering [10], the residual end-of-range (EOR) damage beyond the original a/c interface after regrowth can significantly degrade the device performance due to remarkable reverse bias leakage currents. One solution for reducing the influence from the leakage caused by EOR damage is to widen the spatial separation between the Ge profile maximum and the original a/c interface by using further Si^+ or Ge^{++} implantation to make the a/c interface much deeper [11].

As shown in Fig. 1(a), the RBS spectra reveals that further Ge^{++} implantation of the double-energy Ge^+ and Ge^{++} implantation locates the a/c interface to a deeper position of 390 nm while the single-energy Ge^+ implantation with a corresponding ion energy results in a 265 nm deep a/c interface and a Ge profile maximum at depth of 125 nm. An excellent agreement was found with the GISA-3 spectrum synthesis (solid line) and the experimental spectra (circle). Also the RBS spectra were readily converted to plots of concentration *vs.* depth by the GISA-3 program code, as shown in the inset of the Fig. 1(a). With the GISA-3 treatments the measured Ge distribution is Gaussian in shape with a projected range of 125 nm, FWHM of 43 nm, and the Ge peak content of 8.9%. These results are very well consistent with the TRIM-91 simulations demonstrated in the Fig. 1(b), where the implantation was simulated by only implanting 180 keV Ge ions with a dose of $4.7 \times 10^{16} cm^{-2}$ into a Si

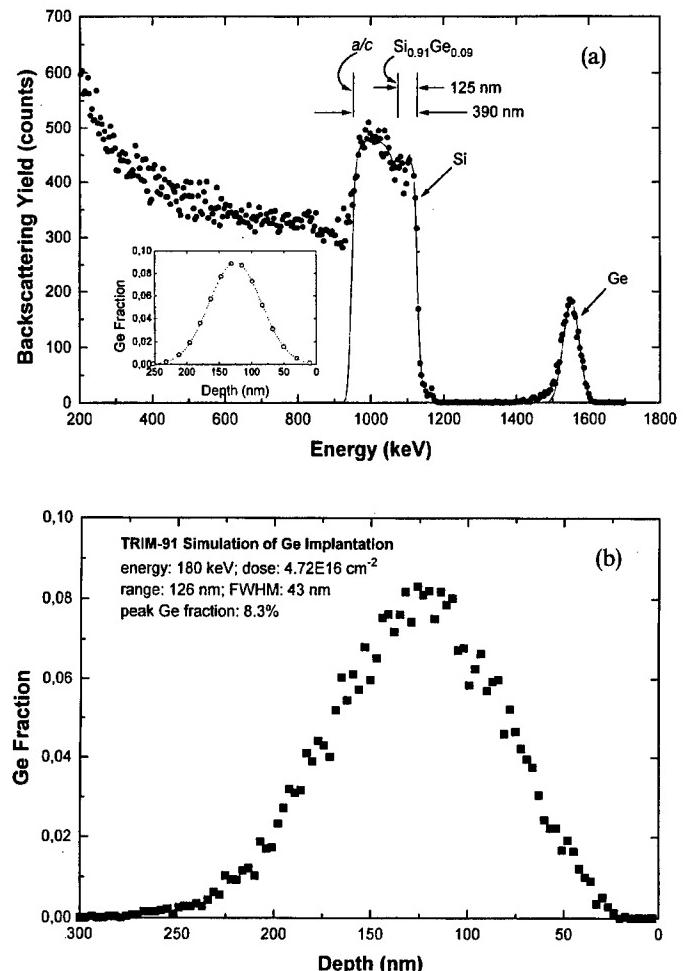


Fig. 1. (a) Aligned <100> channeling spectrum (filled circles) of the sample amorphized by the double-energy Ge^+ and Ge^{++} implantation, while the solid line is a fit with the GISA-3 program. The positions of the Ge profile peak and the a/c interface are shown in the upper part of the figure. Also shown in the inset is the calculated Ge distribution. (b) A Ge depth distribution simulated by full recoil cascade TRIM-91 Monte Carlo calculations of 10 000 primary ions. In calculations the displacement energy and the lattice bonding energy of Si were taken as 15 V and 1 eV, respectively.

substrate. The projected range, the longitudinal straggling and the Ge peak content are 126 nm, 43 nm and 8.3%, respectively. Clearly, this implanted Ge profile is mainly determined by the Ge^+ implantation procedure not by Ge^{++} because of the relatively high dose of Ge^+ . For the double-energy Si^+ and Ge^+ implanted sample, RBS channeling measurements (not shown) confirm that the a/c interface is at depth of about 340 nm while the Ge profile maximum is at depth of 125 nm.

Figure 2 shows annealing properties of the samples implanted by three different methods: Single-energy Ge^+ implantation, double-energy Si^+ and Ge^+ implantation, and double-energy Ge^+ and Ge^{++} implantation. After SPE at 600 °C for 1440 minutes the double-energy Ge^+ and Ge^{++} implanted sample shows the best recrystallization quality among the three implantations with the lowest channeling yield. The minimum channeling yields for Si and Ge in the regrown region are 4.5% and 5.3%, respectively. A slight increase of channeling counts can be found at depth of about 390 nm corresponding to the original a/c interface shown in Fig. 1, which confirms that EOR damage was remained there. For the sample implanted by the double-energy Si^+ and Ge^+ ions the minimum channeling yields for

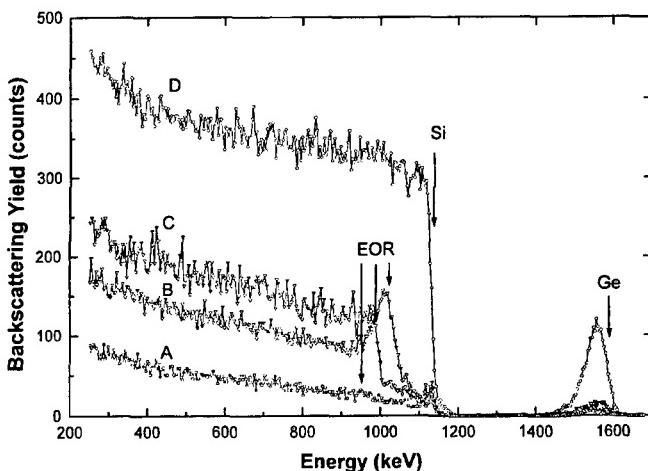


Fig. 2. Aligned $\langle 100 \rangle$ channeling spectra. A: double-energy Ge^+ and Ge^{++} implantation; B: double-energy Si^+ and Ge^+ implantation; C: single-energy Ge^+ implantation; and D: random spectrum of A. All the samples were annealed at 600°C for 1440 minutes.

Si and Ge in the regrown region have been increased to 7.7% and 11.6%, respectively. In addition, the channeling counts of this sample are strongly raised beyond the back-scattering ion energy of 1000 keV corresponding to the original a/c interface, which indicates that much more EOR defects remained beyond the a/c interface after SPE. The channeling spectrum of the sample implanted by the single-energy Ge^+ ions shows the worst recrystallization with respect to the two double-energy implantation methods with the minimum yield for Si in the regrown region larger than 8.0% and even more EOR defects remained near the a/c interface. Therefore, these results lead to the conclusion that the double-energy Ge^+ and Ge^{++} implantation method is optimum with fewer residual defects and wider spatial separation between the Ge profile maximum and the original a/c interface.

In order to understand the regrowth behavior of the sample implanted by double-energy Ge^+ and Ge^{++} ions, this sample was cut into pieces and then individually annealed with different annealing temperature and time combinations. RBS channeling spectra shown in Fig. 3 illustrate the progress of the a/c interface in some annealing conditions. Table I further shows the annealing conditions, the

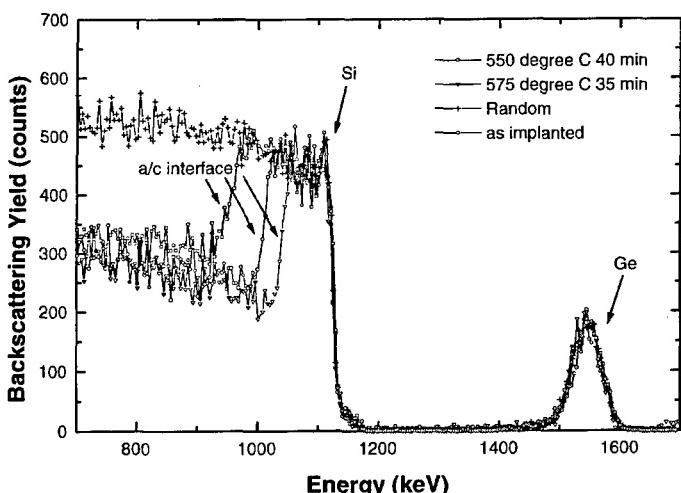


Fig. 3. Aligned $\langle 100 \rangle$ spectra of the double-energy Ge^+ and Ge^{++} ion implanted sample showing the layer regrowth by SPE.

Table I. Regrowth behavior of the double-energy Ge^+ and Ge^{++} implanted sample

SPE	a/c Depth (nm)	Regrowth rate (nm/min)
As-implanted	390.0	0.00
525 °C 100 min	253.8	1.37
550 °C 40 min	262.4	3.22
575 °C 35 min	192.7	5.67
600 °C 9 min	253.8	15.27

RBS measured a/c depth, and the estimated regrowth rate. The regrowth rate were estimated from the amorphous layer thicknesses measured by the RBS channeling spectroscopy. Plotting regrowth rate vs. annealing temperature, shown in Fig. 4, the data points can result in a linear fit corresponding to an activation energy of 1.97 eV, which in agreement with the activation energy of 2.0 eV for $\langle 100 \rangle$ pure Ge, but smaller than the value of 2.4 eV for $\langle 100 \rangle$ pure Si [14]. Regrowth rate values presented in this work were derived only from single measurement points at each temperature interval, a more accurate study on the regrowth kinetics will be planned to determine the regrowth rates by plotting the amorphous depth vs. annealing time.

4. Conclusions

RBS channeling spectroscopy is a very powerful tool for optimising the high-dose Ge implantation into $\langle 100 \rangle$ Si substrates. The calculation results of the GISA-3 computer code, used for analysing the experimental RBS spectra, are in excellent agreement with the TRIM-91 Monte Carlo modelling of the high-dose Ge implantation. The agreement between the results obtained by two different computer codes proves the good quality of the physical modeling. Among the three implantation methods, the double-energy Ge^+ and Ge^{++} implantation is optimum which is confirmed by channeling measurements since it can result in

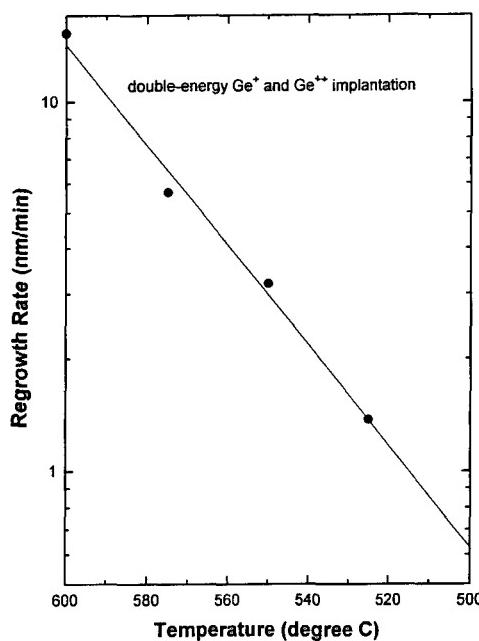


Fig. 4. Plot of regrowth rate versus annealing temperature. The solid line is a linear fit to the data points and corresponds to an activation energy of 1.97 eV.

fewer residual defects, and provides the possibility to widen the spatial separation between the Ge profile maximum and the original a/c interface. Furthermore, the preliminary studies on the regrowth of the double-energy Ge^+ and Ge^{++} implanted sample reveals that the activation energy is on the order of 2 eV.

Acknowledgements

The authors wish to thank J. Salmi and R. Korkeamäki for technical assistance and S. Eränen for helpful discussions.

References

1. Patton, G. *et al.*, IEEE Electron Device Lett. **11**, 171 (1990).
2. Meyerson, B., Scientific American, March, 42 (1994).
3. Winterton, S., Manning, B., Walkey, D. and Tarr, N., Electron. Lett. **27**, 1337 (1991).
4. Paine, D., JOM, February, 55 (1993).
5. Berti, M. *et al.*, J. Mater. Res. **6**, 2120 (1991).
6. Shoji, K., Fukami, A., Nagano, T., Tokuyama, T. and Yang, C., Appl. Phys. Lett. **60**, 451 (1992).
7. Elliman, R. and Wong, W., Nucl. Instr. and Meth. **B80/81**, 768 (1993).
8. Lombardo, S., Priolo, F., Campisano, S. and Lagomarsino, S., Appl. Phys. Lett. **62**, 2335 (1993).
9. Xia, Z. *et al.*, Nucl. Instr. and Meth. B, in press.
10. Grahn, K., Xia, Z., Kuivalainen, P., Karlsteen, M. and Willander, M., Electron. Lett. **12**, 444 (1993).
11. Xia, Z. *et al.*, Appl. Surf. Sci., in press.
12. Jones, K. S., Prussin, S., and Weber, E. R., Appl. Phys. **A45**, 1 (1988).
13. Saarilahti, J and Rauhala, E., Nucl. Instr. and Meth. **B64**, 734 (1992).
14. Lau, S. S., Mayer, J. W. and Tseng, W., in: "Handbook on Semiconductors" (Edited by T. S. Moss and S. P. Keller) (North-Holland Publishing Company, Holland, 1980), p. 531.
15. Feldman, L. C., Mayer, J. W. and Picraux, S. T., "Materials Analysis by Ion Channeling" (Academic Press, New York 1982).
16. Ziegler, J. F. and Biersack, J. P., "The Transport of Ions in Matter" (Version 91.xx), computer code, the code is described in Ref. [17].
17. Ziegler, J. F., Biersack, J. P., and Littmark, U., "The Stopping and Range of Ions in Solids" (Pergamon, New York 1985).

As Capping of MBE-Grown Compound Semiconductors; Novel Opportunities to Interface Science and Device Fabrication

J. K. Grepstad, H. Husby,* A. Borg† and B.-O. Fimland

Department of Physical Electronics, University of Trondheim – NTH, N-7034 Trondheim, Norway

R. W. Bernstein

Department of Microelectronics, SINTEF-SI, N-0314 Oslo, Norway

and

R. Nyholm

Department of Synchrotron Radiation Research, Lund University, Sölvegatan 14, S-223 62 Lund, Sweden

Received May 27, 1994; accepted July 7, 1994

Abstract

In situ condensation of an amorphous cap of the high vapour pressure element (i.e., As, Sb) has been found to provide effective protection of molecular beam epitaxy grown compound semiconductor surfaces against ambient contamination. Most work reported so far relates to arsenic-capped AlGaAs. Detailed investigation with surface sensitive structural (RHEED, LEED) and chemical (XPS) probes confirms that the protective cap is conveniently removed by annealing in ultrahigh vacuum environments at a temperature in excess of ~350 °C. Clean $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ surfaces with different atomic reconstructions and corresponding (Al)Ga : As composition ratios are now routinely prepared by this technique, and thus offers an ideal testing ground for compound semiconductor surface and interface research. Reconstruction-dependent reactivity at metal/GaAs(001) interfaces is demonstrated, using surface sensitive synchrotron radiation photoelectron spectroscopy. Exploiting the protection offered by the As (Sb) cap for device fabrication purposes (e.g., in selective area epitaxy), demands a suitable method of pattern definition in the amorphous arsenic layer. The cap is shown to be chemically stable versus exposure to standard photolithographic processing chemicals, including photoresist, developer, and acetone (the photoresist solvent). However, the temperature required for thermal decapping is grossly inappropriate for photoresist curing. A novel technique of reactive decapping in a beam of hydrogen radicals (H^*) is shown to be effective at room temperature. This innovation makes pattern definition in the As cap compatible with standard photolithography, and test structures with ~5 μm linewidth is demonstrated. Scanning electron micrographs unveil the presence of arsenic cap residues along the photoresist mask edges. Moreover, trace amounts of surface gallium oxide and carbon impurities were found with core-level photoelectron spectroscopy. The technique thus needs further refinement, before being useful in fabrication of compound semiconductor device structures.

1. Introduction

Advances in crystal growth technology, such as molecular beam epitaxy (MBE) and metal-organic chemical vapour deposition (MOCVD), have made possible the fabrication of high-speed electronic and photonic devices which capitalize on the power of heterostructures [1]. If exposed to atmosphere or when handled in compliance with standard substrate preparation procedures, MBE-grown epilayers of, e.g.,

AlGaAs and InGaAs, inevitably form a native surface oxide and contract superficial carbon impurities. Subsequent regrowth on such surfaces leads to free carrier depletion at the interface [2–4], in spite of thermal cleaning in vacuum prior to this growth. As charge transport in heterostructure devices is largely confined to the epilayer interfaces, the device performance is critically dependent on the purity and structural perfection of these interfaces.

Passivation of MBE-grown compound semiconductor surfaces by *in situ* condensation of an arsenic cap, e.g., for ambient transfer from the crystal growth chamber to a second apparatus for further processing or analysis, was first proposed and demonstrated by Kowalczyk *et al.* [5]. Capping and subsequent recovery of $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ epilayer surfaces have since been examined in detail [6, 7]. The arsenic cap is found to provide effective protection of wafers stored in atmosphere for periods up to several months. It is efficiently removed by thermal desorption in ultrahigh vacuum (UHV) environments, at a temperature in excess of ~350 °C. Clean, reconstructed $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ surfaces are now routinely prepared in this manner. Besides $\text{Al}_x\text{Ga}_{1-x}\text{As}$ ($0 \leq x \leq 1$), capping with arsenic was reported to provide effective protection of both InGaAs [8] and ZnSe [9] epilayer surfaces. Moreover, capping with Sb of GaSb was recently reported [10], and underscores the potential of this technique for passivation of compound semiconductor surfaces. In Section 2, we describe the As capping and decapping procedure, as examined with surface sensitive chemical and structural probes.

Core-level photoelectron spectroscopy has long been recognized as a powerful tool for studies of electronic barrier formation at metal-semiconductor junctions, (i.e., the Schottky barrier). Such studies have been largely confined to representative “model systems”, that is to metal depositions on a clean substrate surface with a known crystalline structure, which makes comparison with theoretical description meaningful. Historically, the most popular substrate for investigation of Schottky barrier formation on compound semiconductors has been the (110) cleavage face of (doped) GaAs single-crystals [11]. When cleaved in UHV environments, this non-polar surface exhibits flat-band conditions.

* Present address: Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario M5S 1A4, Canada.

† Present address: Department of Physics, University of Trondheim – NTH, N-7034 Trondheim, Norway.

Subsequent metallization leads to band bending, with the Fermi level shifting towards the center of the gap. This is conveniently monitored by photoemission, as a uniform shift in the substrate core-levels. From the semiconductor bandgap and the measured band bending, the Schottky barrier height is simply evaluated. With the As capping technique at hand, clean, reconstructed epilayer surfaces of the polar (001)-orientation (i.e., the wafer orientation used for compound semiconductor devices) may now be readily prepared for such experiments. In Section 3, we examine the influence of different surface reconstructions on the Schottky barrier formation and interface chemical reactions at Au/GaAs(001) and Al/GaAs(001) junctions, respectively. Motivation came in part from the recent reports of ideal Schottky behaviour for such junctions, if prepared by metallization on decapped MBE-grown surfaces [12, 13], a result in stark contrast to the vast body of previous work [11].

Besides providing easy access to clean, crystalline (wafer) surfaces, passivation with arsenic efficiently suppresses carrier depletion at air-exposed interfaces and in regions of suspended epitaxial (over)growth [14, 15]. Exploiting an As (Sb) cap for surface passivation purposes in the growth and processing of device structures, demands a suitable method of pattern definition in the condensed arsenic layer. Unfortunately, standard photolithography does not lend itself easily to such patterning. The high temperature required for As desorption ($\geq 350^\circ\text{C}$) leads to excessive polymerization of photoresist, which renders subsequent "stripping" virtually impossible. Moreover, the As (Sb) cap must be chemically stable versus exposure to the photolithographic processing chemicals (i.e., photoresist, developer, solvent). The durability of an As cap *vs.* such exposure and the crystalline quality of the epilayer surface after thermal decapping were examined with X-ray photoelectron spectroscopy (XPS) and low-energy electron diffraction (LEED). This data is discussed in Section 4.1.

The problems incurred by thermal decapping through a mask of photoresist on the arsenic cap surface must be worked around, either by finding a substitute for photoresist which may endure curing at $T \geq 350^\circ\text{C}$, or by inventing a method of decapping at reduced temperature, without degrading the quality of the MBE-grown epilayer surface. In

Section 4.2, we demonstrate successful patterning by reactive desorption of arsenic in a beam of hydrogen radicals, through a mask of positive photoresist on the As cap surface.

2. Arsenic capping and decapping of AlGaAs

$\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ epilayers, about 100 nm thick and with different alloy compositions ($0 < x < 1$), were grown on top of 250 nm GaAs buffer layers in a Varian Gen II Modular MBE system. Both semi-insulating and n-type substrates were used, (the latter required for photoemission studies of Schottky barrier formation at metal-GaAs junctions, cf. Section 3). Capping *in situ* of the MBE-grown epilayer is accomplished – most efficiently – with As_2 dimers from an effusion cell equipped with a "cracker". The dimers will stick and form a protective cap at temperatures above room temperature (RT). This is achieved by simply turning off the substrate heater and leaving the wafer to cool in a constant flux of As_2 . In contrast, capping with (less reactive) As_4 tetramers from a standard arsenic oven, demands that the substrate be cooled below room temperature (RT) [5]. Estimated As cap thicknesses in the present study varied between ~ 30 nm and $3\ \mu\text{m}$, dependent on the time elapsed in front of the open "cracker" source at a temperature below $\sim 50^\circ\text{C}$. Figure 1 shows Ga and As LMM Auger sputter depth profiles of an As-capped GaAs sample, allowed to reach a substrate temperature of $\sim 20^\circ\text{C}$ before shutting off the As_2 beam. Ga LMM emission is observed after approximately 9 min of sputtering. The corresponding As cap thickness is estimated at 705 nm [16]. Oxygen KLL emission (not shown in Fig. 1) was found to decrease below noise level after only a few seconds of sputtering, which implies that the thickness of the surface arsenic oxide is less than the depth resolution of the applied profiling technique (i.e., typically a few nm), even after long-term storage of the samples in air.

Figure 2(a)–(d) displays surface-sensitive, high-resolution Ga and As 3d photoelectron spectra of an n-doped GaAs(001) epilayer, capped with arsenic and exposed to air for about one week; as-introduced (a), and after thermal treatment in UHV at $\sim 250^\circ\text{C}$ (b), $\sim 350^\circ\text{C}$ (c), and

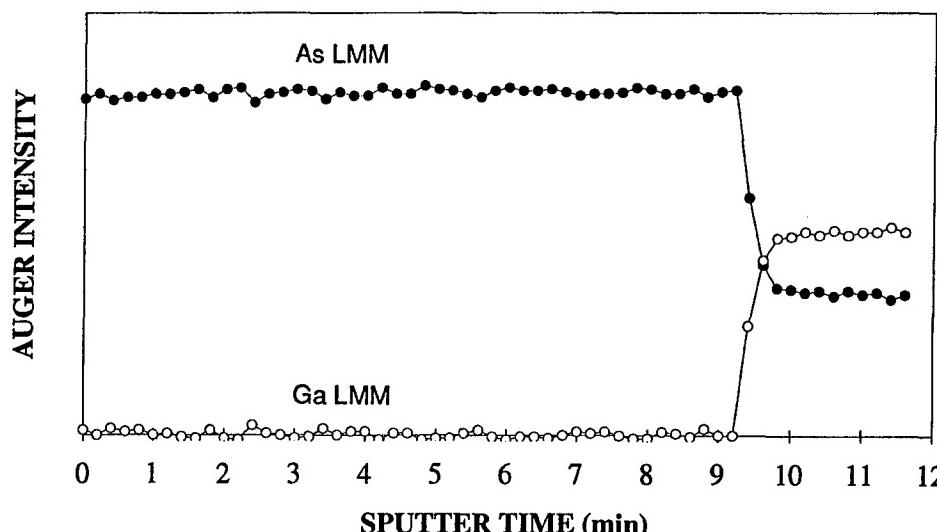


Fig. 1. As LMM and Ga LMM Auger sputter depth profiles of an air-exposed As cap on GaAs(001).

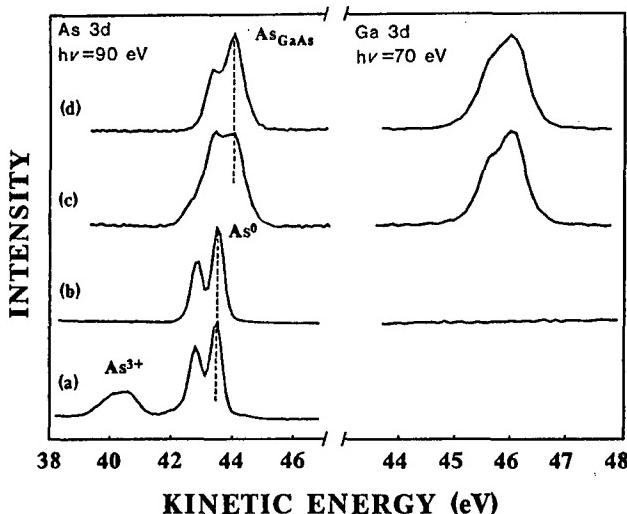


Fig. 2. Surface sensitive As 3d and Ga 3d photoemission spectra of As-capped GaAs(001); (a) exposed to air, (b) after annealing in UHV at $\sim 250^{\circ}\text{C}$, (c) after thermal desorption of the As cap at $\sim 350^{\circ}\text{C}$, and (d) after further annealing at $\sim 450^{\circ}\text{C}$.

$\sim 450^{\circ}\text{C}$ (d), respectively. This data was taken on the toroidal grating monochromator (TGM) beamline 41 at MAX-lab National Laboratory (i.e., the synchrotron radiation facility) in Lund. As 3d_{5/2} emission at three different kinetic energies is distinguished in Fig. 2(a)–(d). The respective shifts in kinetic energy are attributed to As³⁺ of the superficial arsenic oxide ($E_{\text{kin}} = 40.6\text{ eV}$), to As⁰ of the elemental arsenic cap ($E_{\text{kin}} = 43.4\text{ eV}$), and to the covalently bonded arsenic of the GaAs epilayer ($E_{\text{kin}} = 44.0\text{ eV}$) [17]. The As 3d spectrum in Fig. 2(a) suggests that the surface oxide formed upon exposure of the As cap to air is very thin, typically less than 10 Å, even for wafers stored at atmospheric pressure in a sealed plastic box for 3–4 months. The superficial oxide desorbs upon annealing in UHV at $\sim 250^{\circ}\text{C}$, leaving a pure arsenic surface. Thermal desorption of the As cap requires a temperature in excess of $\sim 350^{\circ}\text{C}$. This event is conveniently monitored by a sudden increase in (arsenic) background pressure, and is verified by the onset of Ga 3d photoemission [cf. Fig. 2(c)]. Moreover, the spin-orbit (SO) split As 3d spectrum shifts to kinetic energies typical of the compound semiconductor. The As 3d spectrum in Fig. 2(c) features a minor contribution from As⁰ core-level photoemission, which disappears upon sustained annealing at $\sim 450^{\circ}\text{C}$. Figure 2(d) shows Ga and As 3d spectra of the clean (i.e., non-oxidised), As-terminated GaAs(001) surface.

The recorded As 3d spectrum of the elemental cap [cf. Fig. 2(b)] exhibits a well-resolved SO-split doublet with linewidths of 0.26 eV (FWHM). This is noticeably sharper than the inherent linewidth of GaAs (FWHM = 0.43 eV) [18], and explains in part the loss of energy resolution from Fig. 2(a) and (b) to Fig. 2(c) and (d). Spectral broadening is also contributed by surface shifts in the As 3d core-levels of the decapped epilayer surface. One shifted component pertains to emission from chemisorbed As₂ dimers. This residual surface arsenic desorbs upon annealing at 450 °C, as is evident from the change in spectral lineshape in Fig. 2(c) and (d). (For a detailed discussion of the As 3d spectral lineshape, cf. Ref. [7]). The present results are in close agreement with previous *in situ* photoemission data for MBE-grown GaAs(001) [19]. The thermally decapped epi-

layer surface spectra of Fig. 2(c) and (d) correspond to different surface atomic reconstructions, as discussed below.

Complementary XPS measurements (including data for both Al_xGa_{1-x}As(001) and AlAs(001) epilayer surfaces) were carried out in a Fisons ESCALAB MkII spectrometer in Trondheim. Whereas clean, reconstructed GaAs(001) surfaces were readily obtained after thermal decapping at $\sim 350^{\circ}\text{C}$ and $\sim 450^{\circ}\text{C}$, the preparation (and preservation) of non-oxidised, epilayer surfaces of the alloyed compound (i.e., AlGaAs) proved increasingly difficult with growing AlAs content. Superficial oxide formation, as judged by the O 1s photoemission intensity and a chemically shifted component of the Al 2p spectrum, was observed invariably for AlAs(001) decapped at 450°C , and for Al_{0.5}Ga_{0.5}As(001) after short-term storage in UHV. The elemental As cap is deemed amorphous, from failure to observe a diffraction pattern with RHEED (reflection high-energy electron diffraction) or LEED (low-energy electron diffraction). Electron scattering off an As-capped epilayer surface produces a diffuse halo, only. The GaAs(001) surface is known to reconstruct, i.e., to display a lower symmetry than that produced by a simple termination of the bulk lattice. The actual reconstruction is determined by the surface stoichiometry [20]. Figure 3 shows a comparison of RHEED patterns recorded before capping, i.e., at a substrate temperature of 590 °C and with the epilayer surface exposed to an As₂-flux of BEP = 8×10^{-6} mbar (left), and after thermal decapping at $\sim 500^{\circ}\text{C}$ *in situ*, with a background pressure of 9×10^{-10} mbar (right). Panels (a)–(c) display the RHEED patterns recorded with the electron beam along the [1̄10], [100], and [1̄10] azimuths, respectively. These are the RHEED images characteristic of the As-stabilized (2 × 4)/c(2 × 8)-reconstructed surface, on which AlGaAs epilayer growth commonly proceeds. The close similarity between the two sets of RHEED patterns in Fig. 3 (left and right)

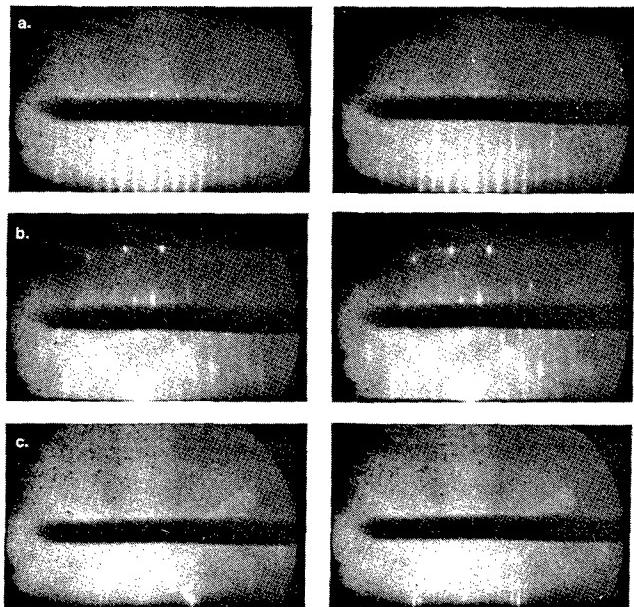


Fig. 3. Comparison of RHEED patterns from MBE-grown GaAs(001), recorded as-grown (left) at $T_{\text{sub}} \approx 590^{\circ}\text{C}$ and exposed to an As₂-flux of BEP = 8×10^{-6} mbar, and after thermal decapping *in situ* (right) at $T_{\text{sub}} \approx 500^{\circ}\text{C}$ and with an (As) background pressure of 9×10^{-10} mbar. The diffraction patterns were obtained with the e-beam along the [1̄10] (a), [100] (b), and [1̄10] (c) azimuths, respectively.

indicates successful recovery of the "as-grown" epilayer surface.

For samples transferred out of the MBE growth chamber and exposed to air, the surface structure after thermal decapping was examined with LEED (in the UHV environments of our ESCALAB). Figure 4 shows representative LEED images recorded after annealing at $\sim 350^\circ\text{C}$ (a) and $\sim 450^\circ\text{C}$ (b). Sharp diffraction patterns characteristic of the $c(4 \times 4)$ -reconstructed surface [Fig. 4(a)] was found for both GaAs(001) and $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ ($x \leq 0.5$). The $c(4 \times 4)$ symmetry pertains to a regular array of chemisorbed As₂ dimers on the As-terminated epilayer surface. The appropriate surface atomic model is depicted in Fig. 5(a), with the $c(4 \times 4)$ unit cell indicated. This model was originally inferred from angle-resolved photoemission data [21], and has later been verified by scanning tunnelling microscopy (STM) [22, 23]. Whereas no LEED pattern of $c(4 \times 4)$ symmetry could be observed for AlAs(001), this reconstruction was verified with RHEED for samples decapped *in situ*. Unlike GaAs(001) and $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ ($x \leq 0.5$), the RHEED pattern of $c(4 \times 4)$ -reconstructed AlAs(001) persists to a substrate temperature of 550°C , (As₂ pressure $\sim 10^{-8}$ mbar).

Thermal decapping of GaAs(001) and $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ ($x \leq 0.5$) at 450°C produces a 1×4 LEED pattern with streaks in the half-order positions along the [110] direction [cf. Fig. 4(b)]. This diffraction image can be explained by coexistent domains of the 2×4 and $c(2 \times 8)$ surface reconstructions, as proposed by Larsen and Chadi [24] and confirmed with STM [23, 25]. The actual surface model (the "missing dimer" model) is displayed in Fig. 5(b), with both

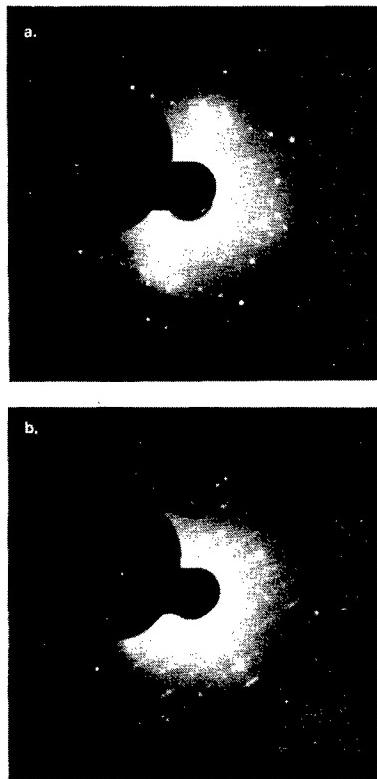


Fig. 4. LEED patterns of clean, reconstructed GaAs(001) epilayers, transferred in air to the photoelectron spectrometer and thermally decapped in UHV at 350°C (a) and 450°C (b), respectively. The diffraction patterns show (a) GaAs(001)- $c(4 \times 4)$ measured at 180 eV, and (b) GaAs(001) - 1×4 with streaks in half-order positions, measured at 189 eV.

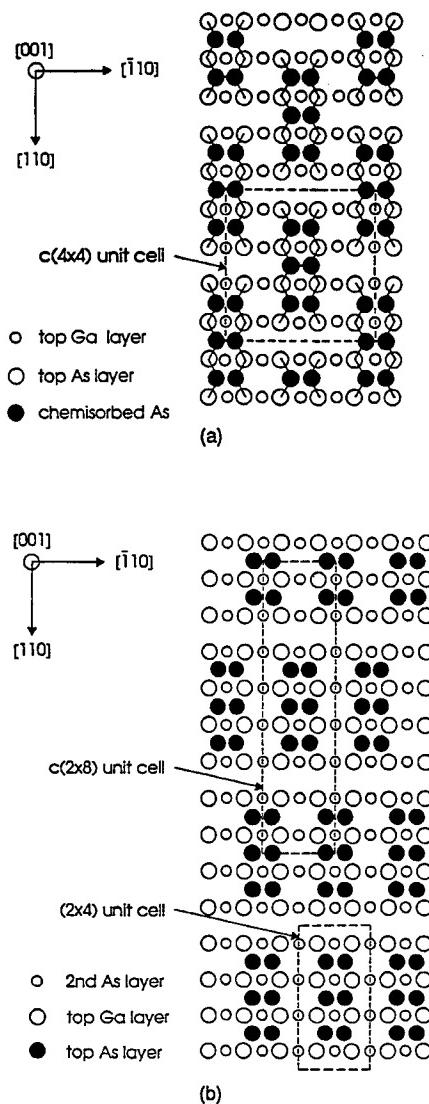


Fig. 5. Surface structure models of the GaAs(001)- $c(4 \times 4)$ (a) and GaAs(001) - $2 \times 4/c(2 \times 8)$ (b) surfaces, (reproduced from Biegelsen *et al.* [22] and from Pashley *et al.* [25]).

unit cells indicated. The failure to observe sharp LEED patterns of surface reconstructed AlAs(001) is attributed to the extreme reactivity of this compound. The decapped AlAs(001) epilayer surface was found to oxidize rapidly, even in the UHV environments of our electron spectrometer.

3. Interface reactivity and electronic barrier formation

Clean GaAs(001) epilayer surfaces of different reconstructions and corresponding Ga : As composition ratios [i.e., As-rich $c(4 \times 4)$, As-terminated $2 \times 4/c(2 \times 8)$, and Ga-terminated 4×6] were prepared by thermal desorption of an arsenic capping layer in the photoemission analysis chamber at MAX-lab. The samples were all n-doped (Si) at $2 \times 10^{18} \text{ cm}^{-3}$. Junctions were formed by *in situ* evaporation of Au and Al. These metals yield the extremes of a wide span of Schottky barrier heights ($\Delta\Phi_b = 0.65 \text{ eV}$) reported in Refs [12, 13] for metal/n-GaAs(001) junctions. (The work function difference between Au and Al amounts to $\Delta\phi = 0.82 \text{ eV}$ [26], for polycrystalline specimens.) Metal deposition rates were calibrated with a quartz crystal monitor.

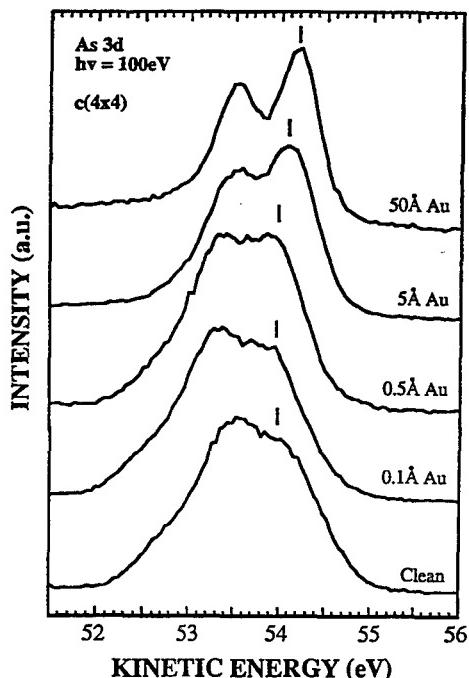


Fig. 6. Evolution of the As 3d core-level photoemission with increasing Au coverage, for depositions on a $c(4 \times 4)$ -reconstructed GaAs(001) surface.

Figure 6 shows the evolution of the As 3d core-level photoemission with increasing Au coverage for the As-rich GaAs(001)- $c(4 \times 4)$ surface, using a photon energy of $h\nu = 100$ eV for maximum surface sensitivity. The recorded spectra are normalized to similar peak heights in order to emphasize changes in the As 3d lineshape. Surface shifted components diminish with increasing Au coverage, and the spectrum sharpens up for depositions in excess of ~ 2 Å. Judging from the recorded peak intensities, the chemisorbed arsenic on the $c(4 \times 4)$ -reconstructed surface segregates on the surface of the Au deposit. From the measured energies of the clean GaAs(001) valence band maximum (VBM) and the Fermi edge of the ~ 50 Å Au film, we may calculate the

initial surface Fermi level (E_f) position with respect to the semiconductor band edges. Subsequent variations in band bending with increasing metal coverage are evaluated from the measured shift in kinetic energy of the As 3d core-levels.

Such data is displayed in Fig. 7 for the three GaAs(001) surface reconstructions examined. In contrast to Chang *et al.* [27], who report near flat-band conditions ($E_f - E_{VBM} = 1.1\text{--}1.3$ eV) for [001]-oriented and vicinal (i.e., 2° -misoriented) GaAs surfaces after thermal desorption of a more than 100 nm thick protective arsenic cap, we find the initial surface Fermi level at $E_f - E_{VBM} \approx 0.5$ eV for the As-rich GaAs(001)- $c(4 \times 4)$ surface and at $E_f - E_{VBM} \approx 0.6$ eV for the As-terminated $-2 \times 4/c(2 \times 8)$ and the Ga-terminated -4×6 surfaces. The additional band bending upon metallization evolves over the first 2 Å of Au deposition and comes to ~ 0.25 eV. This implies a stable Au/GaAs Schottky barrier height of $\Phi_b = 1.05\text{--}1.08$ eV for the two near-stoichiometric surfaces [i.e., GaAs(001) - $2 \times 4/c(2 \times 8)$ and -4×6]. The electronic barrier of the Au/GaAs(001)- $c(4 \times 4)$ junction appears to stabilize at a slightly larger value, $\Phi_b = 1.21 \pm 0.05$ eV, after 10–20 Å of Au deposition. A poor signal/noise ratio in the recorded valence band spectra (including those showing the metal overlayer Fermi edge) denies precise estimates of the Schottky barrier height of the Al/GaAs(001) junctions in this study. No increase (or decrease) in surface band bending was found upon Al deposition to ~ 50 Å coverage. However, our data do indicate a noticeable band bending for the clean, reconstructed substrate surfaces, contrary to the observations of Chang *et al.* [27].

The former reports of ideal Schottky behaviour at metal/GaAs junctions [12, 13] are intriguing, not only in providing unexpected physics, but also with regard to future prospects for device fabrication, (e.g., Schottky diodes and MESFET gate structures with electronic barriers which scale with the metal work function). Correct analysis of this photoemission data is complicated, however. The results of Brillson and co-workers [12, 13] have been called in ques-

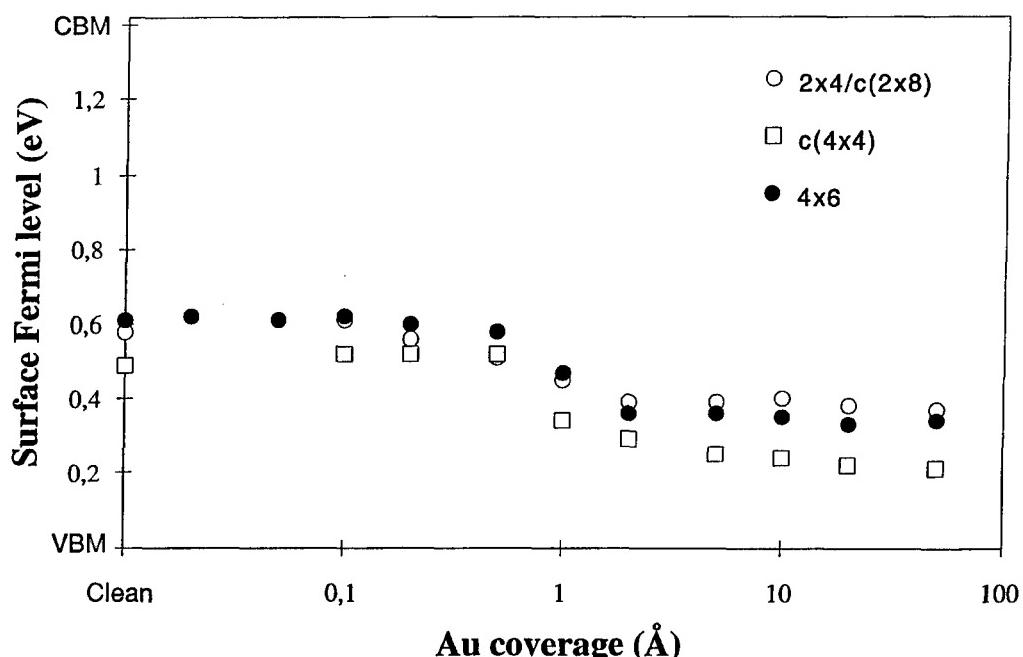


Fig. 7. Surface Fermi level positions *vs.* Au coverage for depositions on the GaAs(001)- $c(4 \times 4)$, $-2 \times 4/c(2 \times 8)$, and -4×6 reconstructed surfaces.

tion, after failure to reproduce their data by groups working at experimental conditions set so as to minimize the influence of a possible surface photovoltage (i.e., ambient temperature, high doping level) [28–30]. Our data for the Au/GaAs(001) – 2 × 4/c(2 × 8) and the Au/GaAs(001) – 4 × 6 junctions is in close agreement with the RT photoemission results of Spindt *et al.* [28], also obtained on MBE-grown n-GaAs(001) epilayers prepared by thermal (As) decapping in UHV. The measured difference in Schottky barrier height between contacts fabricated on the c(4 × 4)- and the 2 × 4/c(2 × 8)-reconstructed GaAs(001) substrates, may be attributed to a difference in work function for the two structurally distinct surfaces, i.e., if we assume near ideal Schottky behaviour for the Au/GaAs(001) junction. Such variations in the work function of MBE-grown GaAs(001) epilayer surfaces have been recently reported by Duszak *et al.* [31].

However, the ideal Schottky model cannot account for the present data in a consistent way. The measured work function of the GaAs(001) – 4 × 6 surface in Ref. [31] is smaller than that of the two other surface reconstructions, and would thus be expected to produce the larger Schottky barrier of the three Au/GaAs(001) junctions investigated, contrary to what is actually observed. The most prominent physical difference between the Au/GaAs(001)-c(4 × 4) junction and those formed on the 2 × 4/c(2 × 8)- and 4 × 6-reconstructed surfaces, is the array of chemisorbed As₂ dimers on the c(4 × 4) surface. It is difficult to see how this excess arsenic would explain the measured difference in Schottky barrier height, however. We note that the measured values for Φ_b are based (in part) on an estimate for the initial surface Fermi level position, with respect to the semiconductor band edges. This estimate for the Au/GaAs(001)-c(4 × 4) junction is seen to differ from that of the two other contacts by approximately the same amount (-0.15 eV) as the final difference in Schottky barrier height (cf. Fig. 7). We feel compelled to caution that the larger electronic barrier for the Au/GaAs(001)-c(4 × 4) junction could come from an error in the estimated initial surface Fermi level position.

Figure 8 shows a comparison of the Ga 3d core-level spectra recorded after $\sim 20\text{ \AA}$ Al deposition on the GaAs(001)-c(4 × 4) and –2 × 4/c(2 × 8) surfaces. Both spectra exhibit a shifted component at higher kinetic energy ($\Delta E_{\text{kin}} \approx 1.0\text{ eV}$). This shift is attributed to dissociated gallium from an (exothermic) exchange reaction with aluminium at the epilayer surface. We note that the reacted Ga 3d peak is far less pronounced for the As-rich GaAs(001)-c(4 × 4) surface, compared to the near-stoichiometric –2 × 4/c(2 × 8) surface. This difference is evident in the recorded photoemission spectra for Al depositions beyond 100 \AA , and suggests that the dissociated Ga diffuses to the aluminium film surface. Pronounced differences in reactivity was previously reported in a photoemission study of Al/2°-misoriented GaAs(001) junctions [27]. The investigators found that the chemical activity correlates with the density of As dangling bonds at steps on the (vicinal) substrate surface, and argue that an As-terminated polar surface thus will be more reactive than its Ga-terminated counterpart, a result in apparent contradiction with the present findings. However, the GaAs(001)-c(4 × 4) surface is effectively passivated by an array of chemisorbed As₂ dimers, contrary to the surfaces investigated in Ref.

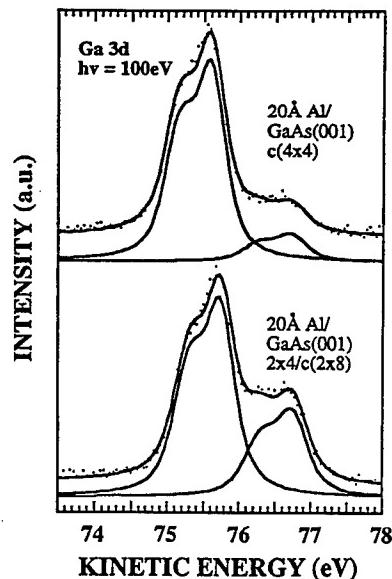


Fig. 8. Comparison of surface sensitive Ga 3d photoemission spectra recorded on the GaAs(001)-c(4 × 4) and the –2 × 4/c(2 × 8) surfaces after 20 Å Al deposition.

[27]. We attribute the reduced chemical activity at the Al/GaAs(001)-c(4 × 4) interface to this As passivation.

4. Photolithographic processing of the arsenic cap

4.1. Durability vs. photochemical exposure

In order to examine the chemical stability of the arsenic cap vs. standard photolithographic processing chemicals, four samples were cut from an As-capped GaAs wafer and exposed ($\sim 15\text{ min}$) to positive photoresist (Shipley 1800), developer (Shipley MF312), acetone (i.e., the photoresist solvent) and NMP (N-methyl-2-pyrrolidone, a common polyimide solvent). Photosensitive polyimide which may be cured at temperatures up to 450°C is now available, and provides an alternative to photoresist for pattern definition. All treatments were terminated with a rinse in deionized water, and the samples were left to dry in air. A fifth sample was taken through complete (maskless) photolithography, including curing of the photoresist at 90°C for 35 min and ultraviolet (UV) exposure for 1 min. Visual inspection with optical microscopy after this processing confirmed complete removal of the photoresist.

Figure 9 shows the As and Ga 2p_{3/2} core-level spectra (recorded with XPS) of the four samples exposed to different photochemicals (b)–(e). The As cap is found to remain intact after the chemical treatments, as is verified by the failure to observe any Ga core-level photoemission from these samples. Spectrum (a) shows corresponding data for an untreated reference sample (i.e., exposed to air only). The two spectral components clearly distinguished in the As 2p_{3/2} spectra of Fig. 9 correspond to emission from a thin, superficial native oxide (As³⁺; $E_b = 1326.4\text{ eV}$) and from the elemental As cap (As⁰; $E_b = 1323.5\text{ eV}$), respectively. The different chemical treatments seem to have no dramatic effect on the arsenic surface oxidation. Maximum As³⁺ core-level peak intensity was found for the sample exposed to NMP [Fig. 9(e)], for which the superficial oxide thickness was estimated at $10\text{--}15\text{ \AA}$ [32]. We note an apparent broadening of the As 2p_{3/2} spectrum after the chemical treatments, which suggests the presence of process-derived

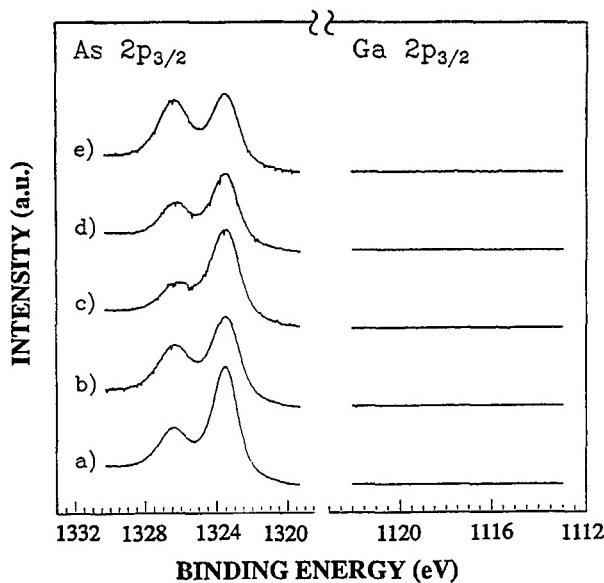


Fig. 9. As 2p_{3/2} and Ga 2p_{3/2} XPS spectra (Al K α excitation) of an As-capped GaAs(001) reference sample (a), and equivalent samples exposed to acetone (b), developer (c), photoresist (d), and NMP (e).

(residual) surface impurities on the capping surface. XPS measurements of the oxygen and carbon core-level photoemission support this conjecture.

The sample which had been exposed to complete (maskless) UV lithography was subsequently decapped by thermal desorption, in order to check for processing-induced surface impurities and structural disorder of the recovered epilayer surface. LEED analysis of this specimen after thermal decapping in UHV unveils the characteristic diffraction pattern of a c(4 × 4)-reconstructed GaAs(001) surface. Core-level photoemission from oxygen and carbon is found to be strongly diminished after decapping. The O 1s emission is reduced to a negligible level and compares with that of a non-processed reference sample, whereas the emission from residual surface carbon clearly exceeds that of the reference surface. Previous studies have established that ambient-derived contaminants desorb entirely upon thermal decapping [7]. Although unclear in what manner, we are inclined to attribute the measured C 1s photoemission to the photochemical processing. From peak intensity

analysis, we find that the residual surface carbon impurities contribute but a small fraction of a monolayer. This explains the observation of a distinct GaAs(001)-c(4 × 4) LEED pattern on the processed and thermally decapped epilayer surface.

4.2. Reactive decapping and pattern definition by H* etching

Attempts at reactive desorption of the As cap at a reduced temperature compared to that of standard thermal decapping, was carried out in the sample preparation chamber of the XPS spectrometer. This effort included, annealing in H₂ gas at 1 atm pressure (i.e., in a high-pressure gas cell incorporated in the fast-entry airlock of the ESCALAB), exposure to molecular hydrogen ions (H₂⁺), and finally, exposure to a thermal beam of hydrogen radicals (H*). Successful decapping at a temperature appreciably lower than that required for thermal desorption, was achieved with the H*-beam, only. This reactive desorption took place at ambient temperature by formation of arsenic hydrides, predominantly AsH₃, and was monitored with a (Fisons SX-200) quadrupole mass spectrometer (QMS). Figure 10 shows a trace of the AsH₃ partial pressure upon H*-etching of Al_{0.3}Ga_{0.7}As(001) with a ~40 nm thick As cap. The start-up procedure involved feeding H₂ gas through the radical beam source (Oxford Applied Research MPD20A), to a measured background pressure of ~4 × 10⁻⁴ mbar. A sudden increase in the AsH₃ partial pressure was observed once the source was powered up (position a). "Ignition" of the hydrogen plasma is seen as a spike (position b). Subsequent r.f.-matching of the connecting circuitry stabilizes the recorded AsH₃ pressure at a slightly reduced level. An order of magnitude increase in the arsine partial pressure is observed once a shutter between the sample and the H* beam is opened (position c), suggesting the onset of aggressive etching of the As cap. The treatment was terminated by closing this shutter, on observation of an abrupt decline in the time trace of Fig. 10 (position d). Subsequent photoemission analysis [cf. Fig. 13(b)] confirmed the complete desorption of the capping layer. (It should be noted that the recorded AsH₃ pressures in Fig. 10 are noticeable lower than their true value, due to a strongly reduced sensitivity of the QMS at this background pressure.) From the recorded time trace in Fig. 10, we estimate the As etch-rate at ~40 nm/min.

Figure 11 shows scanning electron micrographs of two different test-structures "imprinted" in arsenic capping on Al_{0.3}Ga_{0.7}As(001). The patterns were prepared by H* etching at RT through a ~1 μm thick mask of photoresist on the As cap surface. Closeups of the capping edge for the test-structure in Fig. 11(a), are shown in Fig. 12 before (a) and after stripping of the photoresist (b). These micrographs demonstrate that photolithographic patterning of the As cap is feasible, although substantial amounts of residual arsenic remain along the photoresist mask edges. We assume that these As cap residues [excessive in Fig. 11(b)] may be ascribed to premature termination of the H*-beam exposure. A possible strategy for improved edge definition would involve appropriate adjustment of the etching time and of the photoresist and capping layer thicknesses. The jagged shape of the photoresist edge in Fig. 12(a) is attributed to the UV lithography. Inspection of the photomask with an optical microscope showed a rugged pattern edge.

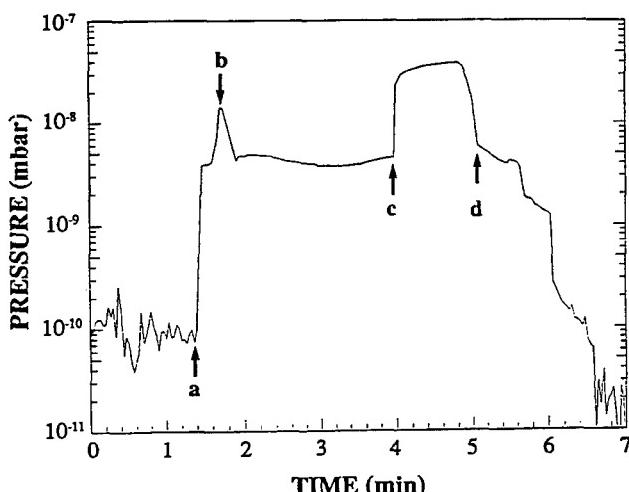


Fig. 10. Time trace of the AsH₃ partial pressure during H* etching of an As-capped Al_{0.3}Ga_{0.7}As(001) epilayer.

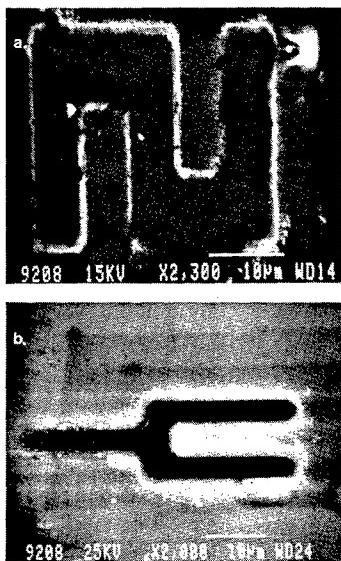


Fig. 11. Scanning electron micrographs of a patterned As cap on $\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}(001)$, for two different test structures (a, b).

Moreover, such profiles typically derive from a poorly tuned exposure time.

Figures 13 and 14 show XPS data for a (maskless) As-capped GaAs sample, as-introduced (a) and after reactive decapping at RT by H^* -beam etching for approximately 2 min (b). The core-level spectra denoted (c) were taken on a thermally decapped, $c(4 \times 4)$ -reconstructed reference surface and are shown for comparison. Effective desorption of the As cap is verified by measurement of Ga $2p_{3/2}$ photoemission and a shift in the As $2p_{3/2}$ binding energy of -0.5 eV in Fig. 14(b). No spectral broadening of the As $2p_{3/2}$ core-level, suggestive of chemisorbed surface arsenic (cf. our discussion of the measured As $3d$ spectra in Section 2), was found in the data taken on H^* -etched specimens. Comparison of the measured As and Ga $2p_{3/2}$ peak intensities with those of the reference spectra (c) suggests a moderate surface arsenic depletion.

The Ga $2p_{3/2}$ spectrum in Fig. 13(b) features a minor contribution at $E_b = 1118.2$ eV, i.e., shifted by 1.1 eV from the

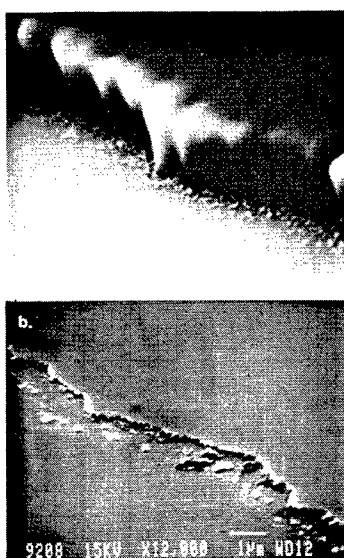


Fig. 12. Scanning electron micrographs showing closeups of the pattern edge [Fig. 11(a)], (a) before and (b) after the final stripping of photoresist.

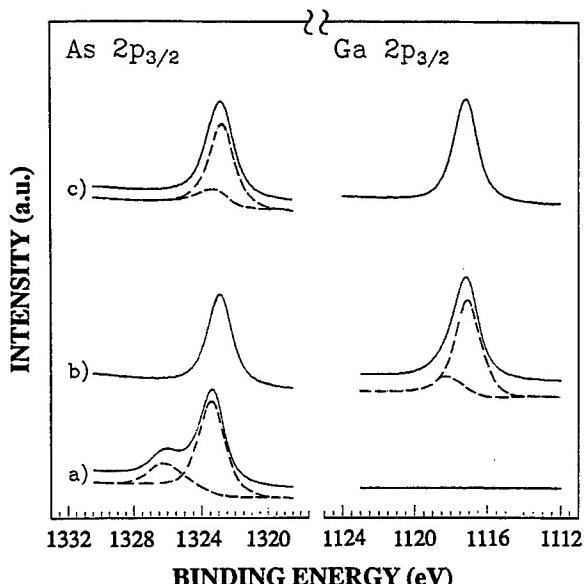


Fig. 13. As $2p_{3/2}$ and Ga $2p_{3/2}$ XPS spectra ($\text{Al K}\alpha$ excitation) of As-capped GaAs, (a) as introduced, and (b) after decapping at RT by H^* etching; (c) shows corresponding data for a thermally decapped $\text{GaAs}(001)-c(4 \times 4)$ reference surface.

bulk GaAs core-level. The shifted gallium core-level photoemission is attributed to a native surface oxide, and this conjecture is corroborated by the measured O $1s$ spectrum [cf. Fig. 14(b)]. The oxygen core-level photoemission exhibits a pronounced assymmetry and may be suitably decomposed in two spectral components. The predominant O $1s$ emission ($E_b = 532.5$ eV) is shifted from that of the native Ga-oxide by about 1.2 eV, and is tentatively attributed to physisorbed H_2O . We note that the presence of hydrogen cannot be ascertained with XPS, and this assignment of the stronger O $1s$ peak in Fig. 14(b) is thus inferred by conjecture. We also note the presence of carbon impurities on the H^* -etched epilayer surface. (The C $1s$ emission in Fig. 14 is superimposed on the high binding energy tail of an As Auger LMM loss satellite). Superficial oxygen (including

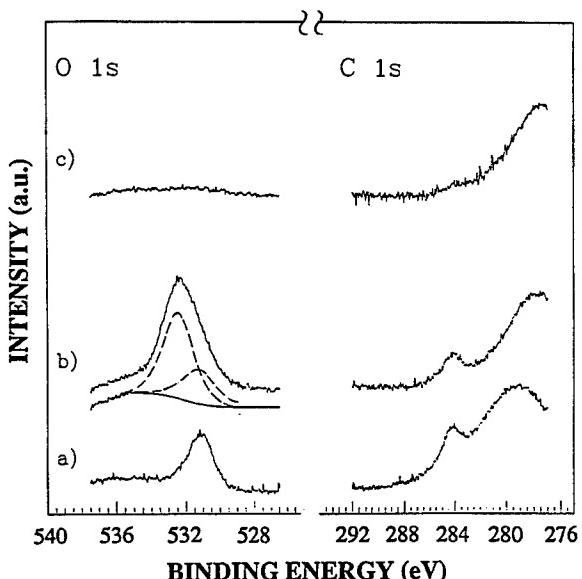


Fig. 14. O $1s$ and C $1s$ XPS spectra ($\text{Al K}\alpha$ excitation) of As-capped GaAs, (a) as introduced, and (b) after decapping at RT by H^* etching; (c) shows corresponding data for a thermally decapped $\text{GaAs}(001)-c(4 \times 4)$ reference surface.

H_2O) and carbon contamination presumably derive from trace amounts of gaseous impurities in the H_2 gas supplied to the radical beam source. Like hydrogen, molecular impurities such as CO_2 and oxygen will dissociate and thus become far more reactive, when passing through the rf-excited plasma of this source. Moreover, surface depletion of As will tend to promote oxidation of the (Al)GaAs epilayer. We finally note that enhanced contamination in the presence of atomic hydrogen (presumably caused by exchange reactions) was previously reported for cleaved GaAs(110) and InP(110) surfaces [33].

LEED analysis of the H^* -etched GaAs(001) surface showed a sharp (1×1) diffraction pattern characteristic of an ordered epilayer, albeit with no distinct surface reconstruction. This is consistent with earlier data of Bringans and Bachrach [34], who reported the LEED pattern of As-terminated GaAs(001) to shift from a $c(4 \times 4)$ reconstruction to (1×1) upon exposure to atomic hydrogen. Moreover, such exposure has been found by several investigators to provide effective cleaning of GaAs wafer surfaces at elevated temperature [35, 36]. Since the As-capped epilayer surface is uncontaminated, we maintain that recovery of a clean, crystalline surface upon etching in a H^* -beam is feasible, even when this etching proceeds at ambient temperature. It is essential, however, that process gas (H_2) of sufficient purity be used and that adequate measures are taken to secure ultraclean processing environments.

5. Conclusion

This study demonstrates that *in situ* condensation of an amorphous As cap in the MBE growth chamber is an adequate technique for protection of $\text{Al}_x\text{Ga}_{1-x}\text{As}(001)$ epilayer surfaces on subsequent exposure to air. Capping in a flux of As_2 dimers eliminates the need for substrate cooling below RT, a prerequisite for condensation of tetrameric arsenic from a standard effusion cell. This technique greatly facilitates safe transportation of samples between different UHV systems and laboratories. The arsenic cap is efficiently removed by thermal desorption in UHV, at a temperature in excess of $\sim 350^\circ\text{C}$. Clean, reconstructed GaAs(001) epilayer surfaces of different Ga:As composition ratios may be prepared by appropriate adjustment of the substrate temperature during this annealing. With AlAs(001), however, surface oxidation is observed almost immediately after decapping. This oxidation adversely affects the surface crystalline order, as judged by the recorded diffraction patterns. Different surface reconstructions were found to have little influence on the electronic barrier formation at (Al)Au/GaAs(001) junctions. However, a distinct difference in reactivity was observed for Al depositions on the GaAs(001)- $c(4 \times 4)$ and $-2 \times 4/c(2 \times 8)$ surfaces.

It was found that the capping remains intact after exposure to the processing chemicals commonly used in UV lithography for integrated circuit and device fabrication. Reactive decapping at ambient temperature was achieved by exposure to a beam of hydrogen radicals. The GaAs(001) epilayer surfaces prepared in this manner were found to be Ga-terminated, with trace amounts of superficial gallium oxide and carbon impurities. Pattern definition in the As cap with $5\text{ }\mu\text{m}$ linewidth was demonstrated, by H^* -beam

etching through a mask of positive photoresist on the capping surface. Further process refinement is needed, however, for definition of sharp pattern edges and in order to eliminate the surface oxygen and carbon contamination. Etching rates and anisotropy of the radical beam treatment should be more closely examined, and the influence of H^* exposure on the electrical properties of the semiconductor epilayer demands attention.

In summary, we find that the As capping technique provides an excellent means of passivation for MBE-grown compound semiconductors. Clean, reconstructed epilayer surfaces are readily prepared by thermal arsenic desorption, and thus offers a valuable testing ground for experiments designed to shed light on fundamental problems in III-V surface and interface physics. Proven durability versus exposure to lithographic processing chemicals and a novel technique for pattern definition in the As cap, suggest that this passivation may also prove useful to device fabrication and in MBE overgrowth on processed multilayer structures.

Acknowledgements

The authors wish to thank the staff at MAX-lab for technical assistance during experiments in Lund. The Norwegian Research Council (NFR-NT) and The Norwegian Telecom Research Department (TF) are gratefully acknowledged for financial support to this work.

References

1. See, for example: Bean, J. C., in: "High-Speed Semiconductor Devices" (Edited by S. M. Sze) (John Wiley, New York 1990), pp. 13–56.
2. Wood, C. E. C. and Joyce, B. A., *J. Appl. Phys.* **49**, 4853 (1978).
3. Kawai, N. J., Wood, C. E. C. and Eastman, L. F., *J. Appl. Phys.* **53**, 6208 (1982).
4. Takamori, A. *et al.*, *Jpn. J. Appl. Phys.* **24**, L414 (1985).
5. Kowalczyk, S. P., Miller, D. L., Waldrop, J. R., Newman, P. G. and Grant, R. W., *J. Vac. Sci. Technol.* **19**, 255 (1981).
6. Schäfer, B.-J. *et al.*, *Surf. Sci.* **20**, 485 (1988).
7. Bernstein, R. W., Borg, A., Husby, H., Finland, B.-O. and Grepstad, J. K., *Appl. Surf. Sci.* **56–58**, 74 (1992).
8. Lau, W. M. *et al.*, *J. Appl. Phys.* **67**, 768 (1990).
9. Farrell, H. H., Tamargo, M. C., Shibli, S. M., Chang, Y. and McNeill, J., *J. Vac. Sci. Technol.* **B9**, 264 (1991).
10. Dumas, M. *et al.*, *Surf. Sci. Lett.* **262**, L91 (1992).
11. See, for example: Rhoderick, E. H. and Williams, R. H., "Metal-Semiconductor Contacts", 2nd edn. (Oxford University Press, Oxford 1988), and references therein.
12. Viturro, R. E. *et al.*, *Appl. Phys. Lett.* **52**, 2052 (1988).
13. Brillson, L. J. *et al.*, *J. Vac. Sci. Technol.* **B6**, 1263 (1988).
14. Kawai, N. J., Nagakawa, T., Kojima, T., Ohta, K. and Kawashima, M., *Electron. Lett.* **20**, 47 (1984).
15. Miller, D. L., Chen, R. T., Elliott, K. and Kowalczyk, S. P., *J. Appl. Phys.* **57**, 1922 (1985).
16. VG SUSPRE software package, VG Ionex; cf. J. P. Biesack, *Nucl. Instr. Meth.* **182/183**, 199 (1981).
17. Mizokawa, Y., Iwasaki, H., Nishitani, R. and Nakamura, S., *J. Electron. Spectrosc. Rel. Phenom.* **14**, 129 (1978).
18. Eastman, D. E., Chiang, T.-C., Heimann, P. and Himpel, F. J., *Phys. Rev. Lett.* **45**, 656 (1980).
19. van der Veen, J. F., Larsen, P. K., Neave, J. H. and Joyce, B. A., *Solid State Commun.* **49**, 659 (1984).
20. See, for example: Joyce, B. A., in: "Molecular Beam Epitaxy and Heterostructures" (NATO ASI Series E: Applied Sciences, No. 87) (Edited by L. L. Chang and K. Ploog) (Nijhoff, Dordrecht 1985), pp. 53–62.
21. Larsen, P. K., Neave, J. H., van der Veen, J. F., Dobson, P. J. and Joyce, B. A., *Phys. Rev. B* **27**, 4966 (1983).
22. Biegelsen, D. K., Bringans, R. D., Northrup, J. E. and Swartz, L.-E., *Phys. Rev. B* **41**, 5701 (1990).

23. Borg, A. *et al.* (unpublished).
24. Larsen, P. K. and Chadi, D. J., Phys. Rev. **B37**, 8282 (1988).
25. Pashley, M. D., Haberern, K. W., Friday, W., Woodall, J. M. and Kirchner, P. D., Phys. Rev. Lett. **60**, 2176 (1988).
26. Michaelson, H. B., J. Appl. Phys. **48**, 4729 (1977).
27. Chang, S. *et al.*, Phys. Rev. Lett. **64**, 2551 (1990); J. Vac. Sci. Technol. **B8**, 1008 (1990).
28. Spindt, C. J. *et al.*, J. Vac. Sci. Technol. **B9**, 2090 (1991).
29. Mao, D. *et al.*, J. Vac. Sci. Technol. **B9**, 2083 (1991).
30. For a discussion of the surface photovoltaic effect, see for example: M. Hecht, Phys. Rev. **B41**, 7918 (1990).
31. Duszak, R., Palmstrøm, C. J., Florez, L. T., Yang, Y.-N. and Weaver, J. H., J. Vac. Sci. Technol. **B10**, 1891 (1992).
32. Bernstein, R. W. and Grepstad, J. K., Surf. Interface Anal. **14**, 109 (1989).
33. Proix, F., Sébenne, C. A., Cherchour, M., M'Hamed, O. and Lacharme, J. P., J. Appl. Phys. **64**, 898 (1988).
34. Bringans, R. D. and Bachrach, R. Z., Solid State Commun. **45**, 83 (1983).
35. Takamori, A., Sugata, S., Asakawa, K., Miyauchi, E. and Hashimoto, H., Jpn. J. Appl. Phys. **26**, L142 (1987).
36. Petit, E. J., Hozay, F. and Moison, J. M., J. Vac. Sci. Technol. **A10**, 2172 (1992); Petit, E. J. and Hozay, F. J. Vac. Sci. Technol. **B12**, 547 (1994).

Process Optimisation and Characterisation of PBT Structures

Sofia Hatzikonstantinidou, Hans-Erik Nilsson*, Christer Fröjd* and C. Sture Petersson

Royal Institute of Technology, Solid State Electronics, Electrum 299, S-164 40 Kista-Stockholm, Sweden

Received June 9, 1994; accepted June 15, 1994

Abstract

The Permeable Base Transistor (PBT) is considered to be a high frequency device with simulated f_T and f_{max} values in the order of 100 GHz. In this work we present several PBT devices in silicon. The fabrication process steps have been developed and optimised in order to meet the demands of a future integration in a standard CMOS processing. Cobalt disilicide is used for the emitter metallization and base metallisation in order to form a good Schottky contact. The important issues of a fabrication process reliability and controllability are discussed in this paper. The process steps had been analysed by standard analysis methods. Electrical (DC) characterisation of the devices has been performed. The obtained results are in a good agreement with the 2D simulations.

1. Introduction

PBTs are considered to operate at frequencies higher than 100 GHz. Several reports concerning the high frequency performance of the PBT have been published. An f_T of 200 GHz for a GaAs PBT has been reported [1]. For a PBT in silicon cut-off frequencies of 120 GHz have been estimated by Monte Carlo calculation [2]. The measured frequency of silicon devices, however, is in the order of 25 GHz [3]. Several techniques have been presented for the realisation of a high frequency PBT device, concerning the geometry [4] and the type of the base [5]. Although extensive work has been done concerning the AC performance of PBTs, very few publications are including DC characterisation [6]. The presented IV characteristics in [6] show a good agreement between measured and simulated data.

This work deals with devices based on the etched geometry using Schottky base diodes. We are aiming at the control of the DC performance by process optimisation. The advantage of the developed process is its compatibility with the already existing CMOS technology and it is based on a self-aligning technique.

2. Device fabrication and process analysis

Silicon n-type, (100) oriented, wafers with a lightly doped epitaxial layer on heavily doped substrate have been used as starting materials. The fabricated devices are mainly distinguished by the properties of the epitaxial layer, in which the channel of the device is formed. The first group of PBTs (in the following text is referred as "PBT 1") was fabricated on 3 μm thick epitaxial layer. The layer is Phosphorus doped with doping level at $4 \cdot 10^{15} \text{ cm}^{-3}$. For the second group (PBT 2) Arsenic doped epitaxial layers have been used. The doping level is $3 \cdot 10^{14} \text{ cm}^{-3}$ and the thickness is 9 μm .

An ion implantation step forms a thin, highly doped layer of 0.2 μm on the top of the epitaxial layer. Arsenic has been used as dopant in all cases. This layer will be the emitter of the devices. Figure 1 shows an SEM micrograph of the developed PBT structure. Figure 2 shows the SIMS doping profile of a typical transistor with respect to the different layers and the model of the device.

As it has been mentioned above, the transistors were developed using the "etched-grooves" technology. This geometry achieves lower device capacitances than the overgrown geometry. The grooves have been formed by dry etching, achieving vertical grid walls. The emitter of the device is formed on the top of the finger-shaped grid, while the bottom is used as the base terminal. The depth of the grooves is a dimensional parameter that varies between different transistors.

Both groups (i.e. PBT 1 and PBT 2) include devices with emitter areas from $(80 \times 80) \mu\text{m}^2$ down to $(40 \times 40) \mu\text{m}^2$. The total transistor area was kept constant and it is $150 \mu\text{m} \times 120 \mu\text{m}$. Table I summarises the common properties and all the variations between the devices and device groups. The grid fingers' width varies from 1.6 μm down to 0.9 μm . The grid fingers are connected by an interconnection bar that is vertical with respect to them. This structure allows us to perform direct DC measurements on the device, therefore any additional lithography step for the formation of contact pads is not required.

After the grid has been formed, a TEOS oxide is deposited in order to achieve the passivation of the walls. The

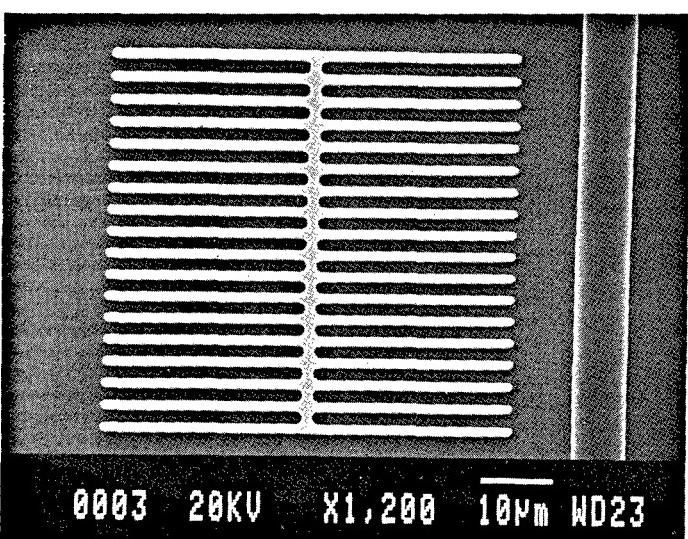


Fig. 1. SEM micrograph showing the top view of a typical PBT device. The device area is $3600 \mu\text{m}^2$ and the finger width is 1.6 μm . The interconnection bar in the middle of the grid connects the grid fingers. Each device is surrounded by a separate frame.

* Also: Mid-Sweden University, School of Science and Engineering, S-851 70, Sundsvall, Sweden
e-mail address: sofiah@ele.kth.se

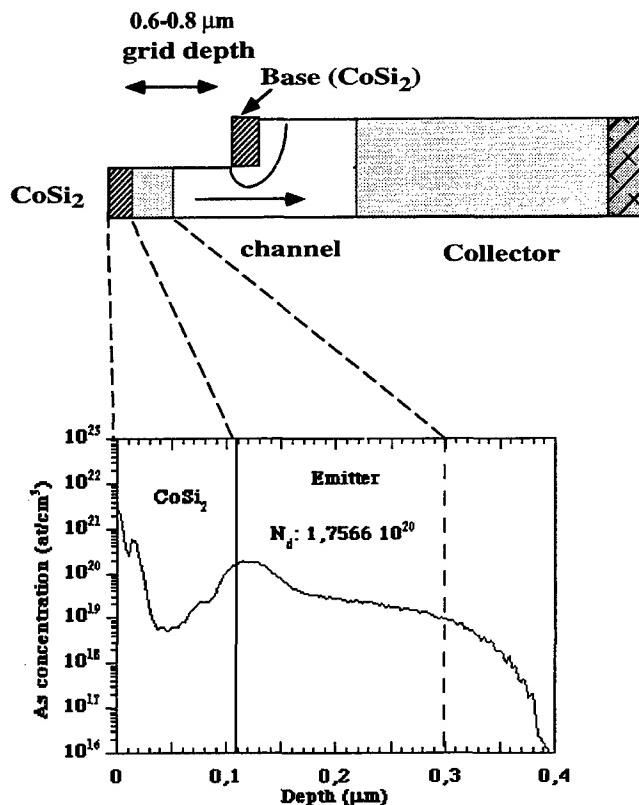


Fig. 2. Model of the developed PBT and schematic representation of the developed PBT structure demonstrating the operation and a qualitative profiling of such device. The SIMS spectra shows a typical As depth profile at the emitter area of the device. The thickness of the silicide has been defined by simultaneous detection of cobalt and silicon intensities. The sample was from a transistor from the PBT 2 group. The emitter layer thickness (0.2 μm) is determined by the As concentration, i.e. at the level of $1 \cdot 10^{19}$ at cm⁻³.

Table I.

	PBT 1	PBT 2a	PBT 2b
Epilayer thickness	5 μm	9 μm	9 μm,
Grid depth	0.8 μm	0.6 μm	0.8 μm
Base thk. CoSi ₂ thk.	850 Å	710 Å	1300 Å
Emitter doping	$1 \cdot 10^{19}$	$1 \cdot 10^{21}$	$1 \cdot 10^{21}$
Channel doping	$4 \cdot 10^{15}$	$3 \cdot 10^{14}$	$3 \cdot 10^{14}$
Φ_b (eV)	0.61	0.62	0.62
Ideality factor η	1.01	1.12	1.03

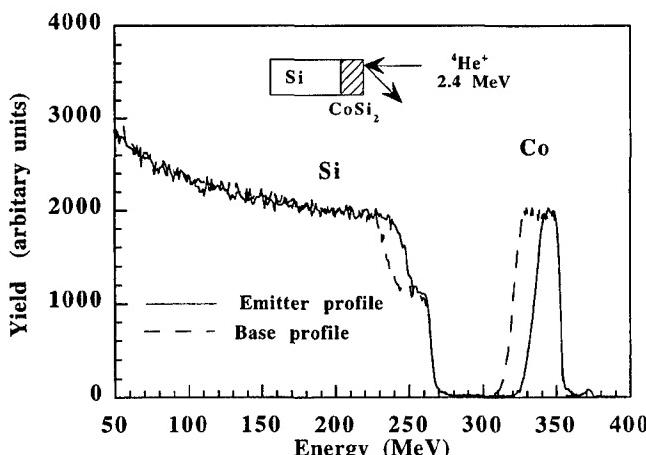


Fig. 3. The RBS profiles about the silicide layer on both emitter and base are presented. It observed that the emitter layer is thinner than the one at the base. The measured thicknesses are 1300 Å and 1180 Å for the base and emitter layers, respectively.

following dry etching step will form the oxide spacers on the grid walls (self-aligned process). This technique is essential for the application of the self aligned silicidation during the following steps.

Cobalt disilicide has been chosen for the metallization of the base and emitter areas, because of its properties like its selective formation, its excellent chemical stability, its low resistivity which make it attractive for applications in ULSI technology and achieves good Schottky contacts to silicon.

The cobalt has been deposited using electron-gun evaporation. The pressure in the evaporation system was better than $8 \cdot 10^{-7}$ torr. Prior to the evaporation the wafers had been dipped in dilute HF-solution to remove the oxide from the silicon surface. The deposited cobalt thickness were 250 Å and 370 Å. The subsequent silicidation was done in an RTA (Rapid Thermal Annealing) system. The wafers were treated at 700 °C for 30 s in nitrogen ambient. The formed silicide layers have been analysed by XRD (X-Ray Diffraction) and RBS (Rutherford Backscattering Spectrometry). The observed orientations of the polycrystalline cobalt disilicide layer from the diffraction spectra are the (111), (220) and (311). The RBS spectra for both the emitter and base areas are presented in Fig. 3. The measured resistivity of all silicide layers was found to be between 15–16 μΩcm. The resistivity of the base is an important parameter for the high frequency behaviour of the PBT and it has to be low [7].

The self-aligned cobalt disilicide was formed only on the top and bottom of the grid. The unreacted cobalt was etched away from the wall spacers by the subsequent selective etching. Thus on top a good ohmic contact was formed, which achieves proper ohmic behaviour of the emitter. On the bottom of the grid a good Schottky base was formed.

The dopant (Arsenic) depth profiles have been analysed by SIMS (Fig. 2). At the same time the concentrations of silicon and cobalt have been detected, thus the profile was analysed with respect to the silicide/silicon interface. The highly doped emitter layer thickness is 0.2 μm. The CoSi₂ layer thickness has been calculated from the RBS and SIMS profiles. The values are listed in Table I.

The silicide layer at the emitter is thinner than the one at the base, as is expected. The silicidation process is retarded on highly doped areas [8] and as the silicidation conditions are not sufficient for the reaction of the whole amount of cobalt, the silicide layer at the emitter is slightly thinner.

Finally, a 1 μm thick aluminium layer was evaporated on the backside of the wafers and alloyed to achieve good ohmic contact to the collector.

3. Electrical characterisation and simulation

The fabricated transistors have been characterised by DC measurements. The probe tips were applied directly on the emitter grid and the base on the large base area. As collector has been used the backside of the wafer.

There is no evidence of bridging, which indicates a reliable and controllable silicidation process. The Schottky base diodes have been characterised by *IV* and *CV* measurements. The results are listed in table I and show the good quality and performance of the Schottky base for all transistors. The leakage currents were below 1.5 μA at 5 V, for all structures.

The first group of devices had a large series resistance due to relatively low (see Table I) doping levels at both the emitter and the collector. The doping level in the other group is significantly higher and achieves a very good ohmic behaviour at both contacts. The measured *IV* curves of a typical transistor is presented in Fig. 4. In order to verify the *IV* characteristics the device structures were simulated in the 2D simulator MEDICI.

The measured current levels for the PBT 1 group were significantly lower than the simulated levels. In this case the doping level on the top of the grid was low (see Table I) and the contact had a rectifying behaviour. Therefore we have to drive the device with the collector-on-top, i.e. this "Schottky contact" will be forward biased. A series resistance at both the emitter and collector contacts, however, has been influencing the characteristics of these devices.

The second group of transistors has current levels and measured *IV* characteristics which agree with the 2D simulations (see Fig. 5).

The threshold voltage is not influenced by series resistances or the quality of emitter and collector contacts. The threshold voltage of the presented PBTs are in good agreement with the theoretical expression (3.1) derived in [10]

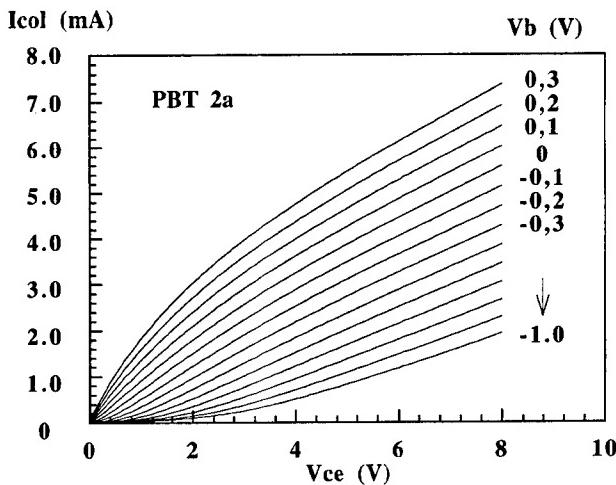


Fig. 4. *IV* characteristics of a transistor from the PBT 2a group (see Table I). The device has an area of $6400 \mu\text{m}^2$ and grid width/spacing of $1.44 \mu\text{m}$.

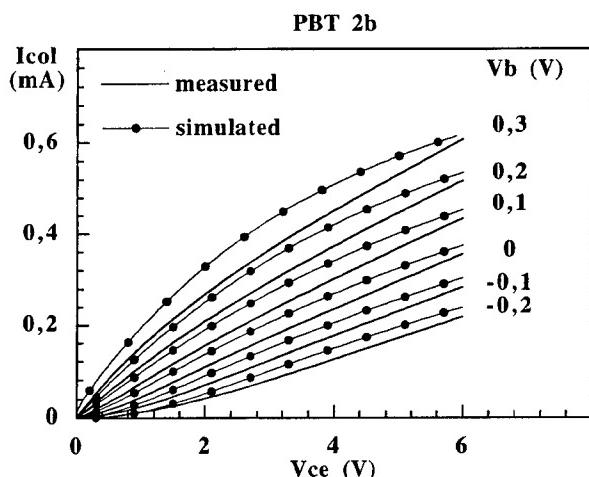


Fig. 5. Comparison between measured *IV* curves and simulation results. The actual transistor has $4000 \mu\text{m}^2$ area and $1.6 \mu\text{m}$ grid width/spacing.

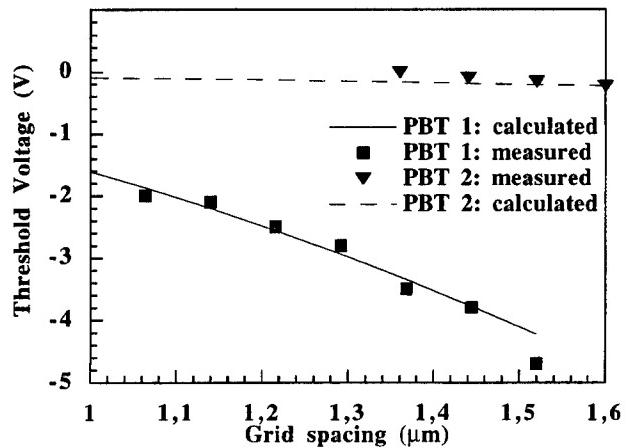


Fig. 6. The comparison between measured and calculated threshold voltages for both transistor groups is presented. The fitting is proofing the quality of the fabricated devices.

and [9] (see Fig. 6).

$$V_t = V_{bi} - \frac{qN_d d_{eff}^2}{8\eta\epsilon_0\epsilon_r} \quad (3.1)$$

where V_{bi} is the contact potential, N_d is the doping level in the channel, d_{eff} is the effective grid spacing extracted from SEM analysis and η is extracted from the two-dimensional simulations which describes the electrostatical behaviour of the grid in two dimensions (lateral and vertical). This parameter is independent of the doping level and the same for all structures.

The good agreement between the 2D simulation model and the measured data allows us to make realistic predictions about the high frequency limits of this structure. The simulated device geometry is based on the minimal dimension achieved by state-of-the-art silicon processes. We considered a grid depth of $0.4 \mu\text{m}$, grid spacing and width of $0.5 \mu\text{m}$ and a constant doping level in the channel in the order of $1 \cdot 10^{17} \text{ cm}^{-3}$. The simulated values for f_{max} and f_T are 58 GHz and 45 GHz, respectively (Fig. 7).

4. Conclusions

Several PBT devices have been fabricated. The developed process steps have been optimised with respect to CMOS compatibility and for future high frequency performance.

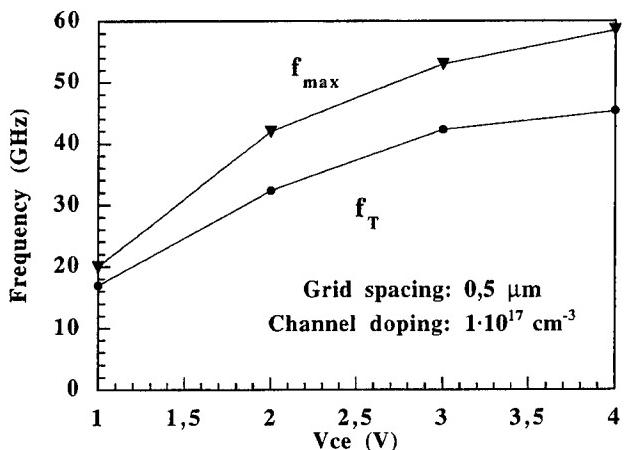


Fig. 7. Simulation results predicting the high frequency behaviour of the developed transistors. The f_{max} and f_T values are at 58,6 GHz and 45,38 GHz, respectively.

The analysis results combined with the electrical measurements prove the controllability and quality of the developed process. The application of cobalt disilicide for the formation of Shottky base contacts has been shown. The DC characterisation of the transistors shows the proper performance of the devices. The controllability of the threshold voltage, considering the dimensions and the channel doping, are qualifying the developed devices. The 2D simulation confirms the operation of the presented PBTs.

A higher channel doping ($1 \cdot 10^{16}$ or $1 \cdot 10^{17} \text{ cm}^{-3}$), a grid size reduction (finger width and spacing) down to $0.5 \mu\text{m}$ and a smaller depth of the grooves can significantly improve the performance of the PBT devices. Furthermore, this process can lead to the fabrication of a device for high frequency performance, since the 2D simulation estimates an $f_T = 45 \text{ GHz}$ and an $f_{\max} = 58 \text{ GHz}$.

Acknowledgement

The authors gratefully thank Margareta Nilsson at the Semiconductor Laboratory, for her proper and careful work during the fabrication process, Margaretha Linnarsson for the SIMS analysis and Nils Lundberg for the

RBS analysis. We wish to appreciate the valuable discussions with our colleagues at the dept. of Solid State Electronics. This work is supported by the Swedish Board for Technical Development.

References

1. Gruhle, A. and Beneking, H., IEEE El. Dev. Let. **11**, 165, (1990).
2. Oshima, T., Nakamura, N., Nakagawa, K., Yamagushi, K. and Miyao, M., Proc. IEDM, 33, (1991).
3. Gruhle, A. et al., Microelectronics Engineering, **15**, 27, (1991).
4. Rathman, D. D., Vojak B. A., Flanders, D. C. and Economou, N. P., 16th Int. Conf. Solid St. Dev. and Mat. 1984, Ext. Abstracts, 305, (1984).
5. van Rijs, F., Oostra, D. J., van Rooij-Mudler, J. M. L. and Timmering, C. E., Proc. ESSDERC 93, 125, (1993).
6. Rathmann, D. D., IEEE trans. Electron Devices **37**, 26, (1990).
7. Sze, S. M., "High-Speed Semiconductor Devices" (John Wiley & Sons, New York 1990), p. 262.
8. Maex, K., Materials Science & Engineering **R11**, 81, (1993).
9. Nilsson, H.-E., Sannemo, U. and Petersson, C. S., Nordic Semiconductor Meeting, **T54**, 159 (1994).
10. Chenevier, P., Kamarinos, G. and Pananakakis, G., Proceedings ESSDERC 93, 121, (1993).

A Flip Chip Process Based on Electroplated Solder Bumps

J. Salonen and J. Salmi

VTT Electronics/EMC, Otakaari 7 B, FIN-02150 Espoo, Finland

Received May 2, 1994; accepted June 15, 1994

Abstract

Compared to wire bonding and TAB, flip chip technology using solder joints offers the highest pin count and packaging density and superior electrical performance. The chips are mounted upside down on the substrate, which can be made of silicon, ceramic, glass or – in some cases – even PCB. The extra processing steps required for chips are the deposition of a suitable thin film metal layer(s) on the standard Al pad and the formation of bumps. Also, the development of new fine line substrate technologies is required to utilize the full potential of the technology. In our bumping process, bump deposition is done by electroplating, which was chosen for its simplicity and economy.

Sputter deposited molybdenum and copper are used as thin film layers between the aluminum pads and the solder bumps. A reason for this choice is that the metals can be selectively etched after bumping using the bumps as a mask, thus circumventing the need for a separate mask for etching the thin film metals. The bumps are electroplated from a binary Pb-Sn bath using a thick liquid photoresist.

An extensively modified commercial flip chip bonder is used for alignment and bonding. Heat assisted tack bonding is used to attach the chips to the substrate, and final reflow joining is done without flux in a vacuum furnace.

1. Introduction

In flip chip technology bare silicon chips are attached upside down on a substrate, which can be another piece of silicon, glass, ceramic or even PCB. IBM's original process, known as C⁴ (for Controlled Collapse Chip Connection) [1], [2] utilized 95%–5% Pb-Sn solder bumps for balls for joining the chip to the substrate. IBM used vacuum evaporation through holes in a metal mask to deposit the solder. This is not, however, very elegant because of the relatively thick layer of solder required. Also, the equipment required is very expensive. An attractive alternative to vacuum evaporation is electroplating. Electroplating is widely used for plating gold bumps for TAB. A suitable binary bath can be used to plate lead and tin simultaneously in a single process step. The baths are compatible with standard liquid photoresists, thus avoiding the cumbersome mechanical masks or the resolution limits of dry resists.

Flip chip processes are similar in many ways to printed circuit board manufacturing and assembly. Fluxes are used extensively used in these processes. In our process, we have purposely avoided the use of fluxes. This can be done if a suitable inert or reducing atmosphere is used in the reflow steps.

Previously published descriptions of flip chip processes using electroplating may be found in references [3] and [4].

2. Test vehicle

A set of masks was designed for process development. The test vehicle comprises a test chip and a corresponding test

substrate. The layouts of both, superimposed on one another, are shown in Fig. 1. The size of the chip is 10,000 μm × 10,000 μm, and the size of the substrate is 10,720 μm × 10,720 μm, and there are 1,228 bumps on the chip. The bumps have been arranged on daisy chains (see Fig. 2) in 19 concentric rings. All bumps in each ring are equidistant from the chip neutral point, and in a bonded

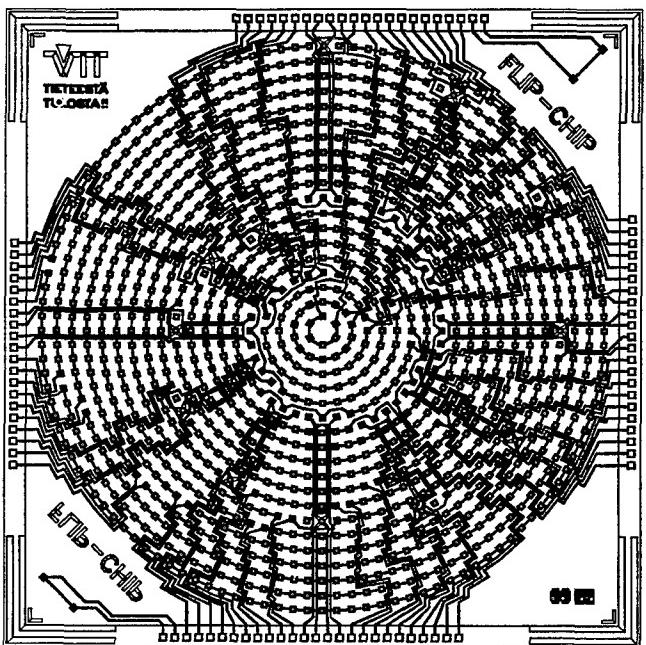


Fig. 1. Layout of test vehicle.

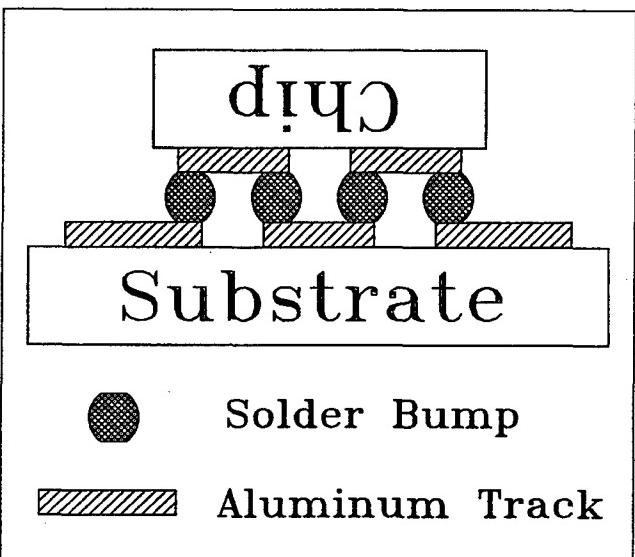


Fig. 2. The daisy chain continuity test structure.

structure are thus subjected to equal stresses due to a possible difference in the thermal expansion coefficients of the substrate and solder materials and/or their unequal temperatures. If the chains were square, instead of circular, the corner bumps would be 41% farther away from the center of the chip, and the corner bumps would also be subject to a mechanical stress also 41% higher. In addition to the daisy chains, there are 4-terminal Kelvin resistor bumps in 16 outermost rings. The nominal diameter of the bumps is 100 µm, and the approximate pitch is 250 µm. The aluminum pads under the bumps are 100 µm × 100 µm squares, and the diameter of the passivation layer opening is 60 µm. The test chips are normally processed on silicon wafers, but any material, compatible with the mask aligner and other process equipment, may be used for the substrates. Only three masks are required for both the chip and substrate processes (or four, if it is desired to use Cr instead Mo in the under bump metallization).

3. Process description

The complete sequence for the test chip process is shown in Fig. 3. 100 mm silicon wafers with 300 nm of SiO₂ on top are used as starting material for test chips. 1 µm of Al is first sputtered on the wafers and patterned with mask M1. This Al layer emulates the topmost metal on a real chip. Next 500 nm of SiO₂ is deposited, using a plasma enhanced CVD process. The oxide layer emulates the passivation layer on a real chip. A second mask, M2, is used to pattern the photoresist for etching the contact holes to Al pads. The process steps hereafter are specific to flip chip technology, and may

be applied to any conventional finished wafer, to convert it from being suitable for wire bonding to being suitable for flip chip bonding.

In IBM nomenclature, the combination of metals used under the bumps is known as BLM (Ball Limiting Metallurgy). A solderable metal (BLM2), such as Ni or Cu, must be used beneath the solder bumps. This makes a solder joint feasible, and additionally limits the solder ball in the reflow phase (the solderable pad is wetted by solder while the passivation layer is not). Also, a suitable barrier/adhesion metal (BLM1) must be used between the solderable metal and the aluminum pads. The standard metal used for this purpose is chromium [5]. In our process, the BLM metals are etched after the electroplating step, using the solder bumps as a mask. The BLM metals, together with the Pb and Sn in the solder, form a complex four-metal system, the choice of which is non-trivial to retain a reasonable selectivity in the etching steps. The apparent reflow properties of the bumps are easily affected, if no fluxes are allowed in reflow. For these reasons, discussed in greater depth in a companion paper [6], we have selected Mo and Cu for BLM1 and BLM2, respectively.

150 nm of Mo is sputtered on wafers, and this is followed by sputtering of 500 nm of Cu. Topologically, for electroplating all bump sites must lie within a simply-connected domain on the wafer surface, which is most easily realized as a blank metal layer covering the whole surface of the wafer. If it is desired to use Cr instead of Mo for BLM1, it must be patterned using an additional mask before the sputtering of BLM2.

The BLM metal layers are next covered with a 20 µm thick photoresist (AZ 4620), which is then patterned with mask M3. After development, there will be 100 µm diameter holes in the photoresist at the sites where the bumps will be plated. The holes in the thick photoresist, as well as the passivation layer openings defined by mask M2, are circular in order to avoid the high local mechanical stress that would inevitably occur at the corners of square shapes. No post-development bake is used for the photoresist to avoid the deformation of the resist profile. The target volume for our bumps is $1.31 \cdot 10^{-12} \text{ m}^3$, which means that the bump shape after plating will be peduncular, as opposed to the sessile shape of the gold bumps normally used for TAB, with only 12% of the total volume in the peduncle. This is of no consequence, as long as the pitch is large enough so that there is no risk of the caps of neighbouring bumps joining into one, as the final bump shape will be defined later in the reflow steps.

A binary bath is used for bump deposition, i.e. both lead and tin are plated simultaneously. We have chosen to use LeaRonal's Solderon M-54 bath, which is based on alcylic sulfonic acid. The singly baked AZ 4620 photoresist withstands this acid very well at room temperature but not at elevated temperatures (+40...50°C). Fig. 4 shows the variation of the lead content of the solder reduced on the cathode as a function of the plating current density for the M-54 bath. This curve was obtained by plating 5 µm of solder on brass cathodes of a size of 36 mm × 46 mm using various currents at room temperature, and by an RBS analysis of the topmost 2 µm in the center of the samples. The peduncular bump shape means that the area of the bump in contact with the bath varies in a ratio of almost

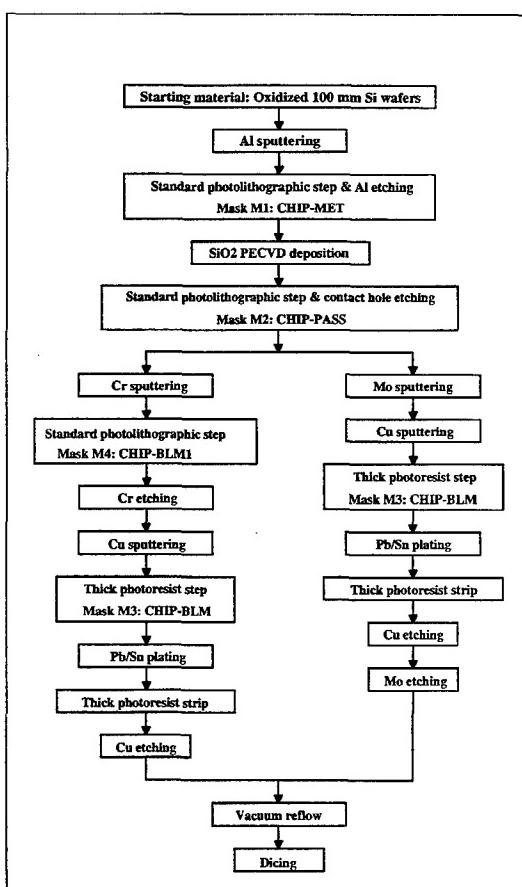


Fig. 3. Test chip process flow. Two alternatives are shown, depending on whether chromium or copper is used for the BLM1 layer.

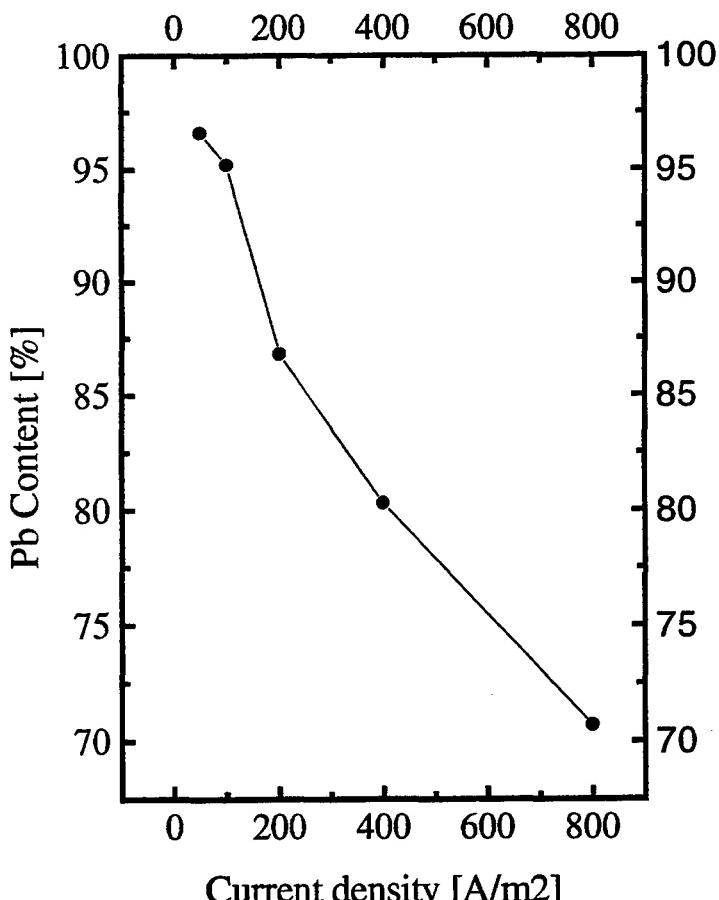


Fig. 4. The measured variation of the lead content (weight per cent) in the solder metal reduced at the cathode as a function of the plating current density. The data is for LeaRonal's Solderon M-54 bath.

1 : 5 during plating. In order to achieve a lead content of 90%, the plating current must be adjusted according to this area variation so that the current density remains constant at 175 A/m^2 . A heuristic model was developed for the bump shape, and it was implemented in a computer program used to control a Keithley 238 current supply. The plating is done on a completely rebuilt AmeriChem Lab-Style plating machine. A piece of platinized titanium mesh is used as a passive anode. A cross section of the bump after plating is shown schematically in Fig. 5.

After plating, the thick photoresist is stripped off. A SEM view of bumps after this step is shown in Fig. 6. The copper layer is etched using an alkaline etchant and the molyb-

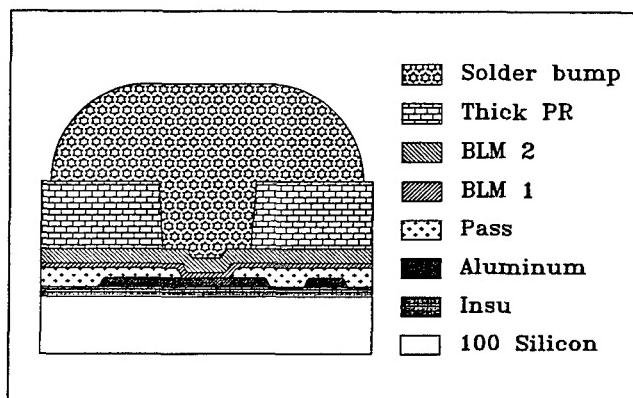


Fig. 5. A schematic cross section (not to scale) of a bump after the plating step.

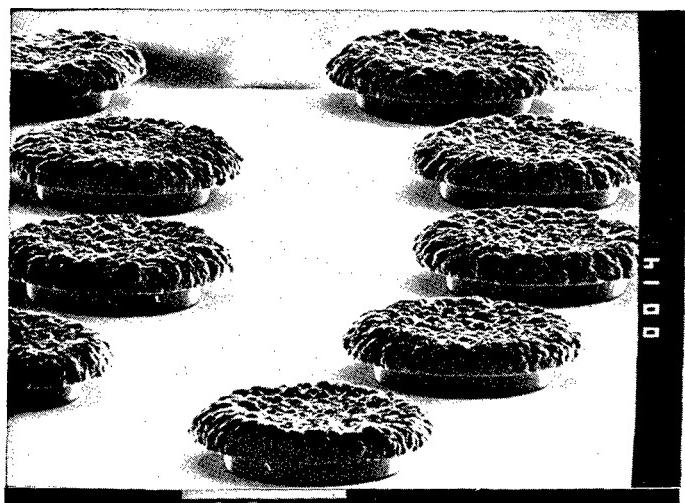


Fig. 6. A SEM view of bumps after thick photoresist strip. The characteristic peduncular shape of the as-grown bumps is clearly visible.

dium using a phosphoric acid etchant [5] using the bumps as a mask. The wafers are then subjected to reflow in a 90%-10% Ar-H_2 ambient at $+400^\circ\text{C}$ at a pressure of 1 mTorr in a homebrew vacuum furnace. In this step, the bumps assume their characteristic spherical shape, limited by the BLM pad underneath. A SEM view of reflowed bumps is shown in Fig. 7.

A somewhat similar three-mask process is used to make the test substrates except there are no bumps on them. Instead, a standard photolithographic step is used to pattern the BLM pads on the substrate side. Additionally, electroless nickel (500 nm) and electroless gold (300 nm), deposited using commercial baths from Shipley, are used to passivate the substrate pads against oxidation which would otherwise occur in the bonding step. The nickel barrier is required to prevent the gold from dissolving into the copper when the substrate lies on the heated stage during alignment. The test chips and substrates are joined together using an extensively modified Laurier TL-271 flip chip bonder (Fig. 8). In this machine, a rather massive substrate holder is heated close to the melting point of the solder. The chip is then aligned with the substrate with the aid of a prism type cube beam splitter and a video camera. During alignment, the distance between the chip and the substrate is about 3 cm. After alignment, the beam splitter is pulled off

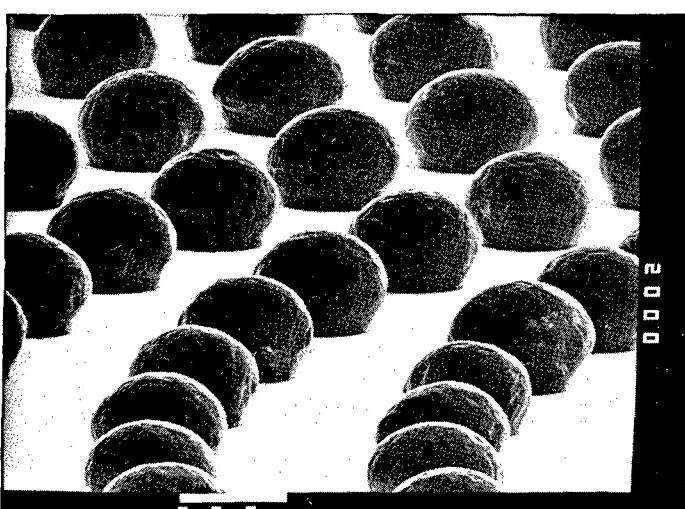


Fig. 7. A SEM view of bumps after reflow.

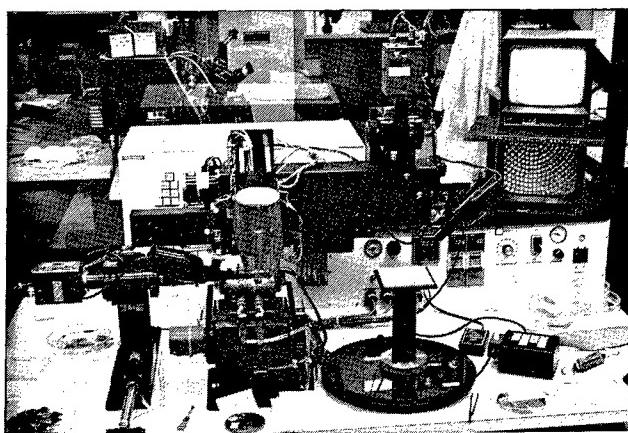


Fig. 8. A general view of the Laurier TL-271 flip chip bonder/aligner. A vacuum collet picks up the chip from the waffle pack (center right) and takes it to above the substrate stage (center left). Alignment is done with the aid of a video camera attached to a beam splitting optical probe (left). The alignment is checked on the video monitor (right).

from between the chip and substrate, and the chip is lowered onto the substrate. The chip is then tack-bonded to the substrate using a pulse of hot nitrogen gas, and the joining is completed in a further reflow step in the vacuum furnace. During this step, the chips float freely on the

molten solder bumps, and the forces resulting from the surface tension of the solder perform the self-alignment unique to this process. No post-reflow clean is necessary, as no fluxes are used. Fig. 9 shows a portion of a test chip bonded to a glass substrate.

4. Summary and conclusions

The outline for a flip chip process based on electroplated solder bumps was presented. The design of a test chip and substrate used for process development was also described. Electroplating provides an economical way to selectively deposit the required amount of solder. Both lead and tin can be plated simultaneously in the desired proportion. A nonstandard feature of the process is the use of molybdenum for the adhesion/barrier metal. Molybdenum was chosen because the etching of more chromium affected the reflow properties of bumps in the flux-free process.

Acknowledgement

The authors wish to thank Mr. J. Saarilahti for performing the RBS analyses.

References

1. Miller, L. F., IBM Journal of Research and Development **13**, 239 (1969).
2. Marcotte, C. M., Koopman, N. G. and Totta, P. A., "Review of Flip Chip Bonding", in "Proceedings of the 2nd ASM International Electronic Materials and Processing Congress" (Edited by W. T. Shieh) (ASM International, Metals Park, Ohio, U.S.A. 1989), p. 73.
3. Maciolek, R. B., "Packaging Very High I/O Chips With TAB/Solder Reflow Technology", in "Proceedings of ASM's 2nd Electronic Packaging: Materials and Packaging Conference" (Edited by J. A. Sartell) (ASM International, Metals Park, Ohio, U.S.A. 1984).
4. Yung, E. K. and Turlik, I., IEEE Trans. Components, Hybrids, and Manufacturing Technology, **14**, 549 (1991).
5. Berry, B. S. and Ames, I., IBM Journal of Research and Development **13**, 286.
6. Salmi, J. and Salonen, J., "Etching of the BLM Layers in a Fluxless Flip Chip Process", paper presented at The 16th Nordic Semiconductor Meeting (Laugarvatn, Iceland 1994).

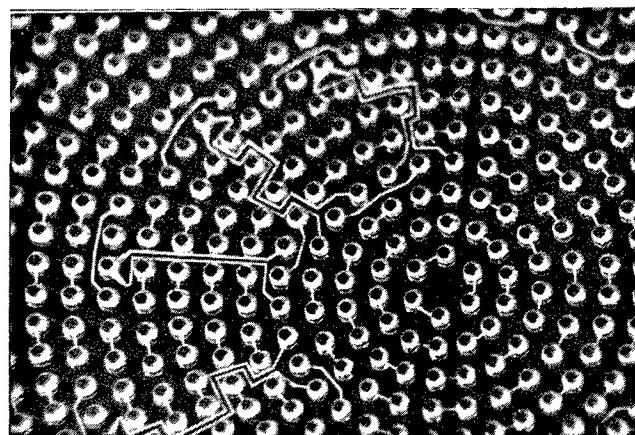


Fig. 9. Test chip bonded to a glass substrate. The bumps are seen through the substrate, the dark circular pads are the 100 µm diameter molybdenum BLM pads on the glass.

Low Temperature Silicon Epitaxy in a Single-Wafer RTP Reactor with Microwave Heating

S.-L. Zhang* and R. Buchta†

* Solid State Electronics, Royal Institute of Technology, Electrum 229, S-164 40 Kista, Stockholm

† Ericsson Components AB, M/AR, S-164 81 Kista, Stockholm

Received June 6, 1994; accepted June 15, 1994

Abstract

Chemical vapour deposition (CVD) of epitaxial silicon was carried out in a single wafer, rapid thermal processing (RTP) reactor chamber. The silicon wafer was heated volumetrically by microwaves. The silicon growth was realised by thermal pyrolysis of SiH₄ at ~700 °C on blank and patterned silicon wafers. The reactor chamber was equipped with a molecular-turbo pump backed by a booster/rotary pump package; the base pressure of the reactor chamber was better than 10⁻⁵ Pa. Silicon deposition was carried out in the pressure range of 1–20 Pa. Prior to deposition, the system was purged with H₂ at 50 Pa, either at 700 °C for 15 min or at 900 °C for 15 sec. A typical epitaxial silicon growth rate was 5 nm/min at 700 °C. The quality of the epitaxial silicon layers, studied by means of Rutherford backscattering spectrometry (RBS), X-ray diffraction (XRD), and cross-sectional scanning electron microscopy (XSEM), was found to be sensitive to a number of deposition parameters, including substrate temperature, gas flow rate and surface cleaning.

1. Introduction

In the silicon integrated circuit (IC) technology, two trends have always been consistent: lateral and vertical downscaling of IC device dimension and increase in wafer size. This evolution places increasingly high demands on thermal budget and calls for more flexible manufacturing technology, which have in turn led to the development of single wafer reactors and rapid thermal processing (RTP) [1]. Device processing with low thermal budget is essential for today's very-large scale integration (VLSI) technology. Low temperature epitaxial growth of silicon is of particular interest for device applications, e.g., high frequency components [2]. Over the past 5 years, we have developed an RTP system that utilises microwave energy to accomplish volumetrical heating of the silicon substrate [3–6]. This is different from the majority of the commercially available RTP systems where wafer heating is realised by surface absorption of infrared light [7]. We have shown that microwave heating of wafers is unique in its uniformity and good temperature control during deposition processes [4] and its inherent microwave plasma facility for *in-situ* surface cleaning [3].

In the microwave heated RTP (MRTP) system, a 100 mm silicon wafer is placed in the electric field minimum of a cylindric reactor chamber which is designed as a resonant cavity preferentially exciting the TE₁₁₂ mode at 2.45 GHz [3, 5, 6]. The microwave energy is inductively coupled to the wafer and the wafer heating is purely resistive. The silicon wafer can be uniformly heated up to 1000 °C within 40 seconds [3, 5]. The reactor is equipped with a molecular-turbo pump backed by a booster/rotary pump package, the base pressure of the process chamber is better than 10⁻⁵ Pa. The process gases of high purity are introduced into the

reactor chamber via respective mass-flow controllers. This system has been utilised for rapid thermal annealing (RTA), as well as for chemical vapour deposition (CVD) [3, 4, 6] and *in-situ* plasma etching [3]. This paper presents new experimental results on silicon epitaxial growth at low temperature (~700 °C) and low pressure (~10 Pa), on blank and patterned silicon wafers. The impact of *in-situ* surface cleaning will be discussed.

2. Experimental

A detailed description of the MRTP system is found elsewhere [3, 6]. Silicon wafers of <100> orientation, 100 mm in diameter and 0.525 mm thick were used in this work. Oxide (SiO₂) layers of either 100 nm or 500 nm thickness were first thermally grown on the wafers. On the wafers with the thick oxide, narrow lines of 1.0 to 5.0 μm width, defined by lithographic technique, were etched into the oxide layer down to the silicon substrate. On the wafers with thin oxide, the oxide on one half of the wafer was stripped off from both sides in an HF solution. Immediately before wafer loading, all the wafers were dipped in diluted HF (1 in 10 H₂O) for 10 sec to remove the native oxide from the exposed silicon surface. The patterned wafers were then rinsed in deionised water, and spin-dried in N₂. The wafers with oxide removed from one half of the wafer surface (blank wafers) were not rinsed in water, leaving a hydrogen passivated silicon surface [8]. After wafer loading, the reactor chamber was evacuated to a base pressure below 10⁻⁵ Pa.

Silicon epitaxial growth was carried out using pure SiH₄ or SiH₄/H₂ mixtures at temperatures from 650 to 710 °C. Before admittance of SiH₄, the reactor chamber was purged with H₂ at a pressure of 50 Pa. During the purge, the wafer was kept at a temperature of 700 or 900 °C. The duration of the purge step was 15 min or 15 sec, respectively. Table I presents the typical experimental conditions for this work. The silicon epitaxy on the blank wafers was characterised by means of X-ray diffraction (XRD) with CuK_α radiation, and Rutherford backscattering spectrometry (RBS) in channelling mode, while the silicon epitaxy on the patterned wafers was studied by cross-sectional scanning electron microscopy (XSEM) in combination with Secco etch [9].

3. Results

3.1. XRD studies on blank wafers

The preliminary verification of epitaxially aligned silicon layer in accordance to the substrate orientation was provided by XRD using a θ-2θ diffractometer. For deposition in pure SiH₄ at a wafer temperature of 690 °C, the XRD results

Table I. Summary of the experimental results

Wafer*	H ₂ purge	Dpto. temp.	SiH ₄ part. pres. and flow rate	Total pressure	Dpto. rate	Film structure†
1	900 °C/50 Pa/15 s	690 °C	6 Pa/0.9 sccm	6 Pa(No H ₂)	4.5 nm/min	align(ml)
2	900 °C/50 Pa/15 s	650 °C	6 Pa/0.9 sccm	6 Pa(No H ₂)	3.9 nm/min	align(h)
3	900 °C/50 Pa/15 s	690 °C	8 Pa/1.2 sccm	8 Pa(No H ₂)	8.4 nm/min	poly
4	900 °C/50 Pa/15 s	710 °C	6 Pa/0.9 sccm	6 Pa(No H ₂)	6.7 nm/min	align(h)
5	700 °C/50 Pa/15 m	710 °C	6 Pa/0.9 sccm	6 Pa(No H ₂)	6.7 nm/min	align(h)
6	700 °C/50 Pa/15 m	710 °C	4 Pa/0.6 sccm	4 Pa(No H ₂)	4.4 nm/min	align(h)
7	700 °C/50 Pa/15 m	710 °C	6 Pa/0.9 sccm	12 Pa(In H ₂)	5.0 nm/min	align(ml)
8	900 °C/50 Pa/15 s	710 °C	6 Pa/0.9 sccm	12 Pa(In H ₂)	5.0 nm/min	align(ml)

* All the wafers were blank, except for wafer #8 which was line-patterned.

† "align" and "poly" denote epitaxial alignment to the substrate orientation and polycrystalline silicon deposition, respectively, on the silicon surface, according to the XRD analysis. "ml" and "h", both in parenthesis, represent the result of surface inspection: mirror-like and hazy, respectively.

(wafer #1) shown in Fig. 1 clearly demonstrate that the silicon deposited on SiO₂ (curve a) is polycrystalline, because of the appearance of the (111), (220), (311), (331) and (422) diffraction peaks, in addition to the strong (400) diffraction peak (at about 2θ = 69°) mainly from the substrate. In contrast, the silicon deposited on the silicon surface (curve b) is epitaxially aligned to the substrate orientation. The deposition rate was 4.5 nm/min.

A reduction of the wafer temperature to 650 °C (wafer #2), and with the other parameters the same as for wafer #1, resulted in a slightly lower deposition rate (3.9 nm/min) as compared to that for wafer #1 (Table I). The silicon deposited on the silicon surface was again found to be aligned to the substrate orientation (curve c, Fig. 1). However, the wafer surface was quite hazy (not mirror-like), revealing the poor quality of the epitaxial layer.

For wafer #3, the deposition was carried out at 690 °C but with a higher SiH₄ flow rate than for wafer #1, in order to obtain a higher deposition rate [10]. The deposition rate increased to 8.4 nm/min, but the silicon deposited *on the silicon surface* was now polycrystalline (curve d, Fig. 1). Increasing the wafer temperature from 690 °C to 710 °C (wafer #4), while keeping the other parameters used for

wafer #1 unchanged, resulted again in alignment of the deposited silicon on the silicon surface according to the XRD results shown in Fig. 2 (curve a). But the surface was still hazy as wafer #2. The deposition rate for this wafer was 50% higher than that for wafer #1 (Table I). The variation of the temperature and time for the *in-situ* surface cleaning in the H₂ atmosphere did not alter the results much (wafer #5), according to the XRD analysis (results not shown). The deposition rates for wafers #2, #4 and #5 were too high to achieve good epitaxial growth at the respective deposition temperatures used.

To reduce the deposition rate, there are many viable approaches. Decreasing the deposition temperature could be a solution, but the results of wafer #2 clearly show that this is not adequate. Reducing the SiH₄ flow rate or decreasing the SiH₄ concentration could also be considered. The former process (wafer #6) did not convert the hazy surface into a mirror-like one, although the deposited silicon was still aligned (curve b, Fig. 2). The latter process with H₂ as carrier gas (wafer #7), however, resulted in a much better growth; besides the epitaxial alignment indicated by curve c in figure 2, the surface was mirror-like. On the same wafer (i.e. #7), the silicon deposition on the SiO₂ surface was polycrystalline (curve d, Fig. 2). The deposition rate was reduced from 6.7 nm/min to 5.0 nm/min.

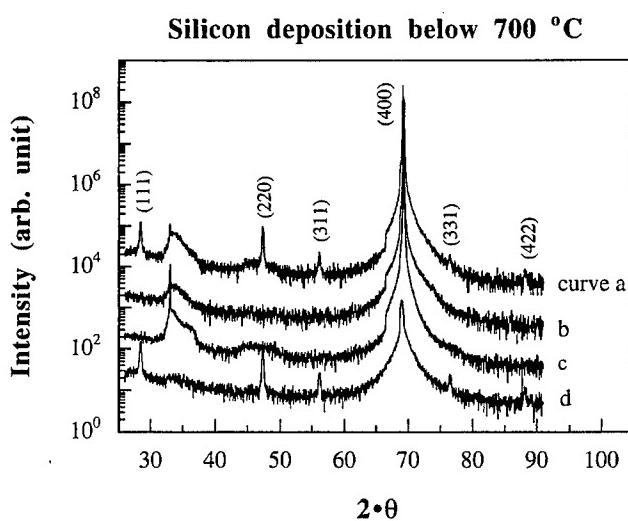


Fig. 1. XRD results of wafers #1 to #3 showing polycrystalline silicon deposition on the SiO₂ surface for wafer #1 (curve a), epitaxially aligned silicon on the silicon surface for wafer #1 (curve b), epitaxially aligned silicon on the silicon surface for wafer #2 (curve c), and polycrystalline silicon deposition on the silicon surface for wafer #3 (curve d). The Miller indices are assigned to the diffraction peaks from the deposited silicon layer.

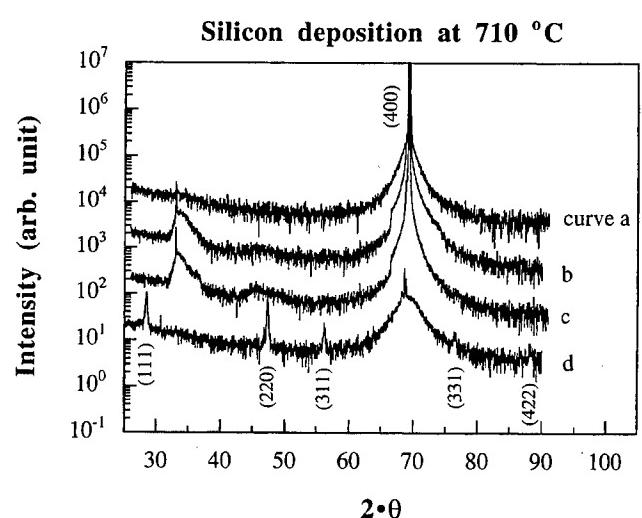


Fig. 2. Similar XRD results of wafers #4, #6 and #7 showing epitaxially aligned silicon on the silicon surface for wafer #4 (curve a), epitaxially aligned silicon on the silicon surface for wafer #6 (curve b), epitaxially aligned silicon on the silicon surface for wafer #7 (curve c), and polycrystalline silicon deposition on the SiO₂ surface for wafer #7 (curve d).

In Fig. 2, the intensity of the diffraction peaks at about 33° (also a diffraction peak from the substrate) and 69° is extremely sensitive to the sample mounting in the diffractometer because the wafers are not exactly in a (100) orientation. The occasional absence of the 33° peak, along with the substantial reduction of the intensity of the 69° peak is attributed to a slight off-orientation of the samples relative to the diffractometer geometry.

3.2 RBS analyses on blank wafers

The XRD results only provide a first indication of the alignment of deposited silicon layers in accordance to the substrate orientation. To investigate the impact of various deposition parameters on the quality of epitaxy, a more sophisticated analysis technique has to be employed. A more rigorous examination of epitaxial growth is provided by RBS in channelling mode. RBS spectra for wafers #1 and #7 are shown in Fig. 3. The observation of a much lower yield of the aligned sample than that for the randomly oriented sample for both wafers provides solid evidence of epitaxial alignment of the deposited silicon layer. The interface between the deposited silicon layer and the substrate is also indicated in Fig. 3. The value of χ_{\min} , characteristic of the quality of epitaxy, is calculated (average of the yields over 5 channels) to be 4.3% and 10.7% for wafers #1 and #7, respectively, indicating a better epitaxy quality for wafer #1 despite a lower deposition temperature used. The best χ_{\min} value for wafer #1 is in fact slightly below 4.0%, if only the yield at one specific channel is used for the calculation. In any case, the χ_{\min} value for wafer #1 is quite close to that reported by Meyerson *et al.* for their epitaxial growth at 750°C and above: i.e. 4.0% [11, 12].

3.3. XSEM studies on patterned wafers

Epitaxial silicon growth was also performed on the patterned wafers with narrow lines of down to $1.0\text{ }\mu\text{m}$ width (wafer #8, table 1). Referring to the RBS results shown in Fig. 3, the *in-situ* H_2 purge procedure used for wafer #1 was chosen, i.e. a short temperature pulse at 900°C for 15 s. For the deposition, the conditions used for wafer #7 were adopted. The deposited silicon layer on the SiO_2 clearly

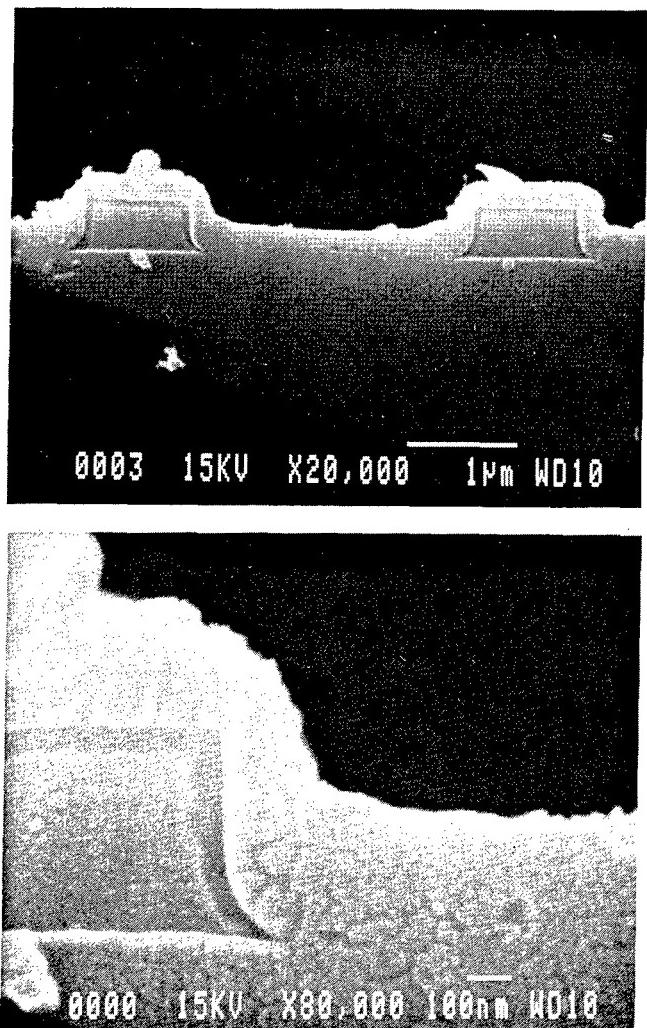


Fig. 4. XSEM studies of wafer #8 (Table I), after the Secco's etch for 20 sec, providing further evidence on (a) polycrystalline silicon deposition on SiO_2 , and epitaxial silicon growth on silicon, and (b) the silicon deposition on the side-walls of the SiO_2 windows is also polycrystalline under the deposition conditions used.

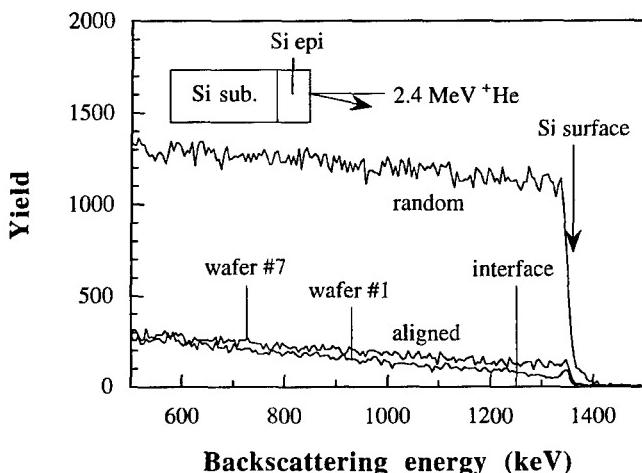


Fig. 3. RBS results verifying the epitaxial silicon growth on the silicon substrate for wafers #1 and #7. The value for χ_{\min} is calculated to be 4.3% for wafer #1 and 10.7% for wafer #7.

shows typical column-like polycrystalline structures, as revealed by XSEM (Fig. 4). This is different from the layer deposited on the silicon surface in the SiO_2 windows, where the deposited silicon is essentially free of such column-like structures (4a). The Secco etch (of 20 sec) has removed some of the polycrystalline silicon on top of SiO_2 , while the epitaxial silicon grown on the substrate surface is almost intact. The XSEM picture with higher magnification (4b) further shows that the silicon deposited on the side-walls of the SiO_2 layer is also polycrystalline, pointing to the predominance of polycrystalline silicon growth in such areas under the present deposition conditions. It is also seen that the epitaxial growth on the silicon surface contains local defects at the interface, which can be attributed to the presence of local surface contamination such as trace native oxide. To remove such surface contamination, a more effective *in-situ* surface cleaning technique is needed.

4. Discussion and conclusions

The quality of the silicon epitaxy is sensitive to a number of deposition parameters: the substrate surface condition, the deposition rate and the wafer temperature. The surface

cleaning prior to the deposition is crucial and any contamination on the surface leads to failure of epitaxial growth. The approach taken in this work was a short etch in diluted HF prior to wafer loading. Pre-baking in H₂ was always performed *in-situ* in order to remove the trace native oxide remaining on the substrate surface and to passivate the surface with hydrogen atoms. Such baking in H₂ above 800 °C has been shown to be very effective for the removal of native oxide [13]. The major difference between wafer #1 and wafers #6 and #7 was the different *in-situ* surface cleaning procedures utilised; for these wafers, the deposition rates were quite comparable. Thus, the experimental results (Fig. 3) seem to confirm that the temperature used for this *in-situ* surface cleaning has to be sufficiently high in order to be effective.

It has been shown that the deposition temperature and deposition rate are the other two major parameters that determine the quality of silicon epitaxy. The deposition temperature is of prime importance, because it determines the thermal migration of the adsorbed silicon atoms on the substrate surface. For a given deposition rate, there exists a temperature threshold below which the epitaxy becomes very difficult. If the deposition rate is too high, the adsorbed silicon atoms would not have sufficient time to migrate before they are buried by newly adsorbed silicon atoms. This is indeed the case for wafer #2. In general, the transition from polycrystalline silicon growth to epitaxial silicon growth took place in the temperature range of 700–750 °C [8].

In summary, a more sophisticated *in-situ* surface cleaning method is required to insure a surface adequate for high quality silicon epitaxy. *In-situ* surface cleaning methods involving the use of different reactive gas species, such as GeH₄, HF and NF₃, or using microwave plasmas have been examined [1]. Because of the inherent microwave plasma facility with the MRTP system [3], *in-situ* surface cleaning with microwave plasmas is the approach we are

considering. Under proper control of the surface condition, silicon epitaxy of high quality can be obtained below 700 °C.

Acknowledgements

We are grateful to Prof. F. d'Heurle for valuable comments, to G. Böling and G. Solås for their skilful microwave engineering, to J. Svennebrink for valuable discussions on microwave engineering, to N. Lundberg and M. Östling for RBS analysis, and to M. Jargelius for cross-sectional SEM work. Part of this work was financially supported by the Ångström-Consortium at Uppsala University.

References

1. Moslehi, M. M. *et al.*, IEEE Trans. Electron Devices **ED-39**, 4 (1992).
2. Ashburn, P., Shafi, Z. A., Post, I. R. C. and Gregory, H. J., in "Proceedings of the 23rd European Solid State Device Research Conference" (Edited by J. Borel, P. Gentil, J. P. Noblanc, A. Nouailhat and M. Verdone) (Grenoble, France, 13–16 September 1993), pp. 301–308, and references therein.
3. Buchta, R., Böling, G., Sellberg, F. and Sigurd, D., 22nd European Microwave Conf. Proc. 1, 34 (1992).
4. Buchta, R., Zhang, S.-L., Sigurd, D. and Lindgren, K., *Appl. Phys. Lett.* **62**, 3153 (1993).
5. Buchta, R., Sigurd, D. and Svennebrink, S., in "1st International Rapid Thermal Processing Conference" (Edited by R. B. Fair and B. Lojek) (The Phoenician, Scottsdale, Arizona, September 8–10, 1993), pp. 359–362.
6. Zhang, S.-L., Buchta, R. and Sigurd, D., *Thin Solid Films*, **246**, 151 (1994).
7. Rapid Thermal Processing 1993 Volume 1 – Companies, Products, Markets, Salzer Technology Enterprises, Inc., 1993.
8. Meyerson, B. S., Himpsel, F. J. and Uram, K. J., *Appl. Phys. Lett.* **57**, 1034 (1990).
9. Schimmel, D. G., *J. Electrochem. Soc.* **123**, 734 (1976).
10. Regolini, J. L., Bensahel, D., Scheid, E. and Mercier, J., *Appl. Phys. Lett.* **54**, 658 (1989).
11. Meyerson, B. S., *Appl. Phys. Lett.* **48**, 797 (1986).
12. Meyerson, B. S., Ganin, E., Smith, D. A. and Nguyen, T. N., *J. Electrochem. Soc.* **133**, 1232 (1986).
13. Hsieh, T. Y., Jung, K. H., Kwong, D. L., Koschmieder, T. H. and Thompson, J. C., *J. Electrochem. Soc.* **139**, 1971 (1992).

The Au/Si(111) System Studied by Optical Second-Harmonic Generation

K. Pedersen^a and P. Morgen^b

^a Institute of Physics, Aalborg University, Pontoppidanstræde 103, DK-9220 Aalborg Øst, Denmark

^b Fysisk Institut, Odense University, Campusvej 55, DK-5230 Odense M, Denmark

Received May 2, 1994; accepted June 14, 1994

Abstract

The room temperature growth of Au on Si(111) 7×7 has been followed continuously with second-harmonic generation, and at selected coverages with low energy electron diffraction and Auger electron spectroscopy, during the evaporation of the first 15 monolayers. A fast decay of the second-harmonic signal in the submonolayer regime is associated with disordering of the 7×7 structure. This is followed by a local maximum at about 1 monolayer coverage. The origin of this maximum is not yet fully understood. In the range from 2 to 5 monolayers, the signal variations reflect the formation of Au islands with areas of free Si in between. Above 5 monolayers second-harmonic generation shows increasing order in the system which is ascribed to the creation of an ordered Au/Si interface, starting when a closed Au layer is formed. Annealing of the room temperature deposited Au layer leads to interface ordering even at low temperatures where no superstructure is observable with low energy electron diffraction.

1. Introduction

The Au/Si(111) system has been investigated by a variety of surface sensitive techniques with the purpose of studying the room temperature growth of thin films and the formation of ordered structures on annealed surfaces. Scanning tunnelling microscopy (STM) studies have contributed substantially to the understanding of ordered structures [1, 2]. A number of experiments have supported a model where the Au, after a critical coverage, grows as an intermixed Au-Si layer on top of the substrate [3, 4]. However, recent experiments [5, 6] have suggested that room temperature growth leads to an abrupt interface between metallic Au and bulk Si.

Second-harmonic generation (SHG) has previously been used on the Au/Si system. McGilp [7] studied the Au/Si interface at selected coverages in the 2 to 12 monolayer (ML) range using a sample cut 4° from a (111) plane. Evidence for ordered terraces at high coverages was obtained. Au deposition at high substrate temperatures has earlier been monitored by SHG and variations in surface reconstruction with Au coverage were identified [8]. In the present work the growth of the first 15 ML of Au on Si(111) 7×7 is studied by SHG. With this technique it is possible to follow the growth process continuously during evaporation and, even more importantly, to study the Au/Si interface below the growing overlayer, since the attenuation of light through a thin Au layer is very small.

2. Theory

Optical SHG is a result of a nonlinear polarization $P(2\omega)$ induced by an applied laser field $E(\omega)$:

$$P_i = \chi_{ijk}^{(2)} E_j(\omega) E_k(\omega), \quad (1)$$

where $\chi^{(2)}$ is the second order susceptibility tensor. In the bulk of centrosymmetric media such as Si, all elements of $\chi^{(2)}$ vanish. Only surface and interface regions with broken inversion symmetry will contribute to SHG. The non-vanishing elements of $\chi^{(2)}$ will then reflect the macroscopic symmetry of the interface [9]. Also spatially nonlocal terms of higher order than the dipole term in eq. (1) can contribute. At the surface, these terms can be included in an effective surface dipole susceptibility [10]. For a surface with $3m$ symmetry the SH fields of interest are given by

$$E_{pp}(2\omega) = a[f_1 \chi_{zxx} + f_2 \chi_{xxz} + f_3 \chi_{zzz} + f_4 \chi_{xxx} \cos(3\phi)] \times E_p(\omega)^2, \quad (2)$$

$$E_{sp}(2\omega) = a[f_5 \chi_{zxx} + f_6 \chi_{xxx} \cos(3\phi)] E_s(\omega)^2, \quad (3)$$

$$E_{ps}(2\omega) = a f_7 \chi_{xxx} \sin(3\phi) E_p(\omega)^2, \quad (4)$$

where $E_m(2\omega)$ indicates the n-polarized SH field for an m -polarized pump field and the f_i are Fresnel factors. The x -axis is parallel to the [211] direction and z is along the surface normal. In the present experiments, the angle ϕ between a [211] direction and the plane of incidence was chosen to 30° . This makes it possible to separate χ_{xxx} and χ_{zxx} by proper choice of polarization combinations. Terms involving χ_{xxx} vary with ϕ and show the symmetry of the system. Thus, for $\phi = 30^\circ$, the p_1 to s_2 polarization combination (p -polarized pump to s -polarized SH) can be used to monitor order in the system.

The surface sensitivity of SHG is particularly high if the nonlinear excitation involves localized surface states. This has been shown to be the case for the 7×7 reconstructed Si(111) surface [11]. With 1064 nm (1.17 eV) pump light a near resonant excitation from filled rest atom states situated 0.9 eV below the Fermi edge to unoccupied adatom states gives the dominating contribution to χ_{xxx} .

3. Experimental

The silicon samples were 1 mm thick n-type, P-doped wafers with a resistivity of 4.5 ohm-cm. Clean 7×7 surfaces were obtained by direct resistive heating to about 1100°C . The UHV system was designed for optical measurements with a number of closely spaced windows and equipped for low energy electron diffraction (LEED) and Auger electron spectroscopy (AES). Gold was evaporated from a Au wire heated by a tungsten coil. The evaporation rate (about 0.7 \AA/min) was measured by a quartz oscillator and the pressure during evaporation stayed below $1.4 \times 10^{-10} \text{ mbar}$. The SHG measurements were performed with a Q-switched Nd:YAG laser delivering 10 ns pulses at 20 Hz with a wavelength of 1064 nm. The energy density was

kept below 20 mJ/cm^2 . A photomultiplier tube connected to gated electronics was used to detect the SH signals.

4. Results and discussion

Figure 1 shows the intensities of SHG during Au deposition on Si(111) 7×7 for different polarization combinations of the pump and SH light using an angle of incidence of 35° . All three configurations show a fast decay of the SH signal in the submonolayer regime. This is consistent with observations with LEED showing a gradual decay of the 7×7 structure and thus the loss of the dangling bonds states responsible for the relatively large signals from the 7×7 surface. Also common to the three polarization configurations is the local maxima in the signals at about 1 ML coverage. This was not accompanied by any observable features in the LEED pattern. Since a maximum is also present in the p_1 to s_2 signal some ordering of the system takes place around 1 ML. This might be a result of changes in the size and distribution of Au islands revealing larger areas of free Si surface. The ordering may not be strong enough to show up in a qualitative LEED investigation. However, further experiments are necessary before a more precise interpretation of these local maxima can be made.

In the range from 2 to 5 ML no order is detected in the system with SHG. At 3 ML it was still possible to see traces of 7×7 reconstruction with LEED which shows that the Au forms islands. In this coverage range the s_1 to p_2 signal, described by χ_{zxx} , grows up to a maximum. The χ_{zxx} tensor element describes a polarization normal to the surface generated by a pump field along the surface. A microscopically flat surface with full translation symmetry is not expected to give any appreciable SHG in the s_1 to p_2 case. Corrugated surfaces, like the 7×7 surface and surfaces with islands, on the other hand, can lead to a nonvanishing χ_{zxx} . The growth of the s_1 to p_2 signal in the 2 to 5 ML range is therefore taken as a sign of island formation.

Photoemission spectroscopy [5] has shown that the formation of a silicon skin layer saturates at a coverage of 4–5 ML. At higher coverages a continuous Au layer grows between the skin layer and the substrate. In the present experiments, no LEED pattern was detected for coverages above 5 ML and the Si(LVV) peak had evolved into two clearly separated peaks. This shows that no free Si surface is left. The increasing order detected by SHG is therefore a result of ordering of the Au/Si interface. At 10 ML the interface ordering is completed and no further change in the p_1 to s_2 signal is observed. The increase in order coincides with the decreasing χ_{zxx} , the tensor element which we have attributed to the existence of islands. Thus, the ordering of the interface takes place when a closed Au layer is formed and the areas with free Si surface between the islands disappear.

SHG can be used to detect the symmetry of the interface by rotating the sample about the surface normal. At normal incidence, where only χ_{xxx} contribute to SHG, the symmetry can also be detected by simultaneously rotating the polarization directions of the pump and SH light. In Fig. 2 the rotational anisotropy of a surface covered by 28 \AA Au is compared to that of the clean 7×7 surface. Clearly, both systems show 3-fold symmetry with some minor 2-fold symmetry superimposed. The 2-fold symmetry of the 7×7 surface is a result of defects in the surface created by the direct heating process [12]. This defect structure is not likely to be carried over to the Au/vacuum interface during the growth process. The observation of the 2-fold symmetry on both clean and Au covered surfaces is therefore taken as further evidence of the existence of an ordered Au/Si interface.

The p_1 to p_2 polarization configuration involves the three tensor elements χ_{xxx} , χ_{xxz} and χ_{zzz} . At coverages above 10 ML the s_1 to p_2 signal shows that χ_{zxx} vanish. The remaining two elements are those which usually dominate SHG from metals [13]. In particular χ_{zzz} can be large for clean metal surfaces [14]. It is seen that the p_1 to p_2 signal

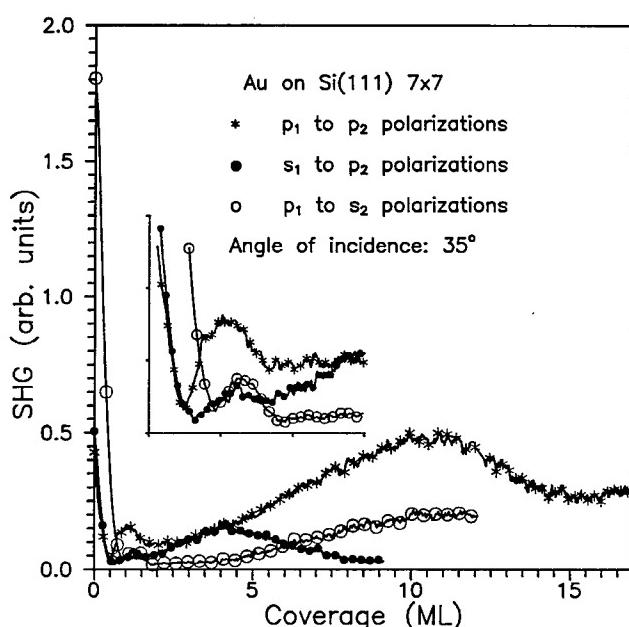


Fig. 1. SH intensities as a function of Au coverage measured with three different combinations of incident and detected light polarizations. The insert shows an expansion of the coverage range up to 3 ML.

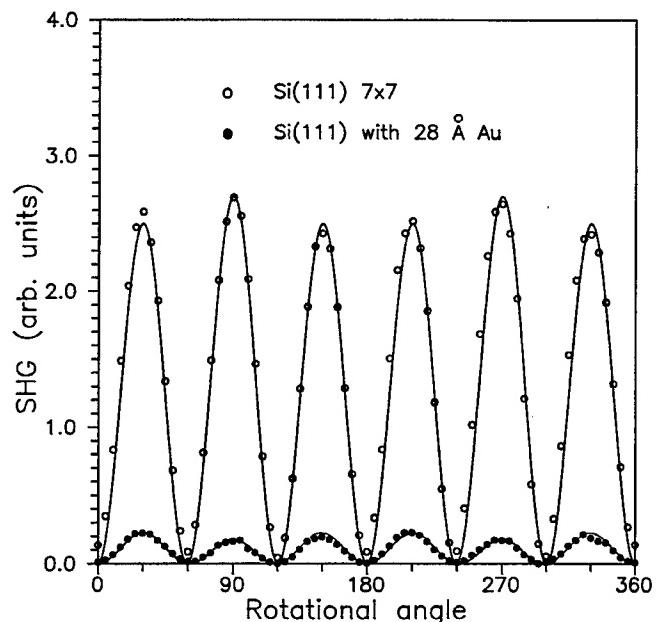


Fig. 2. SH intensity as a function of the polarization direction of the laser light at normal incidence for a clean 7×7 surface and for a surface with 28 \AA Au. The angle is measured relative to the $[2\bar{1}\bar{1}]$ direction and the SH light is polarized perpendicular to the incident light.

increase at coverages above 5 ML as the creation of the Si skin layer has saturated and the Au layer grows below. Above 10 ML the signal saturates and then decrease to a minimum. The origin of the variations at high coverages is not understood at present and it remains to be seen if the signal shows oscillations with coverage which could indicate quantum well effects. Clearly, from the point of view of SHG, the formation of the Au/Si system has not reached a saturation at 15 ML.

Annealing of room temperature deposited Au layers leads to ordering of the surface into structures which depend on the initial thickness of the Au layer. Figure 3 shows the variations of the SH signals with annealing temperature for a surface initially covered by 24 Å Au. The measurements have been made with the sample at room temperature between each annealing at subsequently higher temperatures. The sample was held at each temperature until no further change in SHG was observed. This procedure was chosen instead of direct monitoring during annealing because the signals showed a temperature dependence.

Ordering of the system takes place even at low temperatures. Below 400 °C no order was detected by LEED and the ordering detected by SHG must take place at the Au/Si interface. Also the p_1 to p_2 signal increases while χ_{zxx} remains zero. According to the previous discussion of tensor elements, χ_{zxx} is expected to originate from the Au surface. Thus, no changes take place at the Au/vacuum interface. This is consistent with recent AES experiments [6] showing that the Au and Si peaks are constant up to 370 °C. Above this temperature the Si peak increase as Au diffuse into the

bulk. In the present experiments ordering into a 6×6 structure was observed after annealing at 400 °C. This coincides with the appearance of χ_{zxx} and a decrease in the p_1 to p_2 signal. The order in the system seen with SHG improve with annealing temperature and at 700 °C the surface transforms into a $(\sqrt{3} \times \sqrt{3})R30^\circ$ structure. Finally, above 1150 °C, the Au is removed from the surface and a sharp 7×7 LEED pattern appears. The ratios between the signals from the 7×7 and $\sqrt{3} \times \sqrt{3}$ surfaces found here agree with results obtained by O'Mahony *et al.* [8] by following SHG during deposition of up to 1.2 ML Au on a substrate kept at 800 °C.

5. Summary and conclusions

By using appropriate choices of input and output polarization directions the order of the Au/Si system, and in particular the interface, have been studied by SHG as a function of coverage and annealing temperature. A strong coverage dependence with two orders of magnitude signal variation during the evaporation of the first monolayer shows the decay of the 7×7 structure. In the coverage range from 5 to 10 ML the growth of an ordered interface is detected. This suggests that the Au/Si interface is abrupt in agreement with recent conclusions drawn from experiments with electron spectroscopies [5, 6]. Annealing leads to improved interface order, even for low temperatures where no ordered structure is observed by LEED. The onset of surface reconstruction and the formation of 6×6 and $\sqrt{3} \times \sqrt{3}$ structures are clearly identified in the SH signals.

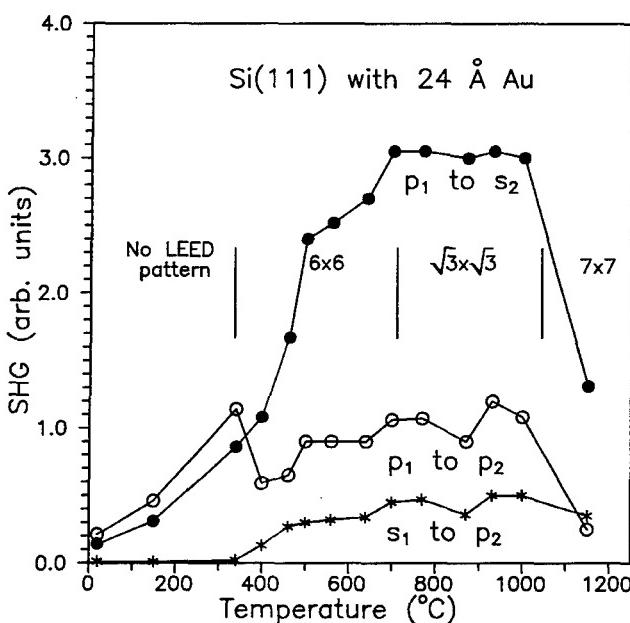


Fig. 3. SH intensities as a function of annealing temperature for three different combinations of incident and detected light polarizations shown together with the dominating LEED patterns obtained in different temperature intervals.

References

- Baski, A. A., Nogami, J. and Quate, C. F., Phys. Rev. **B41**, 10247 (1990).
- Shibata, A., Kimura, Y. and Takayanagi, K., Surf. Sci. **273**, L430 (1992).
- Braicovich, L. *et al.*, Phys. Rev. **B20**, 5131 (1979).
- Perfetti, P. *et al.*, Phys. Rev. **B26**, 1115 (1982).
- Yeh, J.-J. *et al.*, Phys. Rev. Lett. **70**, 3768 (1993).
- Morgen, P., Simonsen, A. C. and Pedersen, K., Proc. First Int. Symp. on Control of Semiconductor Interfaces, in press.
- McGilp, J. F., J. Vac. Sci. Technol. **A5**, 1442 (1987).
- O'Mahony, J. D., Kelly, P. V. and McGilp, J. F., Appl. Surf. Sci. **56-58**, 449 (1992).
- Sipe, J. E., Moss, D. J. and van Driel, H. M., Phys. Rev. **B35**, 1129 (1987).
- Guyot-Sionnest, P., Chen, W. and Shen, Y. R., Phys. Rev. **B33**, 8254 (1986).
- Kelly, P. V., Tang, Z.-R., Woolf, D. A., Williams, R. H. and McGilp, J. F., Surf. Sci. **251/252**, 87 (1991).
- Nakamura, K., Yamamoto, H., Domen, K., Shimizu, H. and Hirose, C., Jpn. J. Appl. Phys. **32**, 3584 (1993).
- Liebsch, A. and Schaich, W. L., Phys. Rev. **B40**, 5401 (1989).
- Murphy, R., Yeganeh, M., Song, K. J. and Plummer, E. W., Phys. Rev. Lett. **63**, 318 (1989).

Fabrication of Nanostructures Using MBE and MOVPE

J. Ahopelto

Electronic Materials and Components, VTT Electronics, Technical Research Centre of Finland, Otakaari 7B, SF-02150 Espoo, Finland

H. K. Lipsanen, M. Sopanen, T. Koljonen and T. Tuomi

Optoelectronics Laboratory, Faculty of Electrical Engineering, Helsinki University of Technology, Otakaari 1, SF-02150 Espoo, Finland

and

V. M. Airaksinen, J. Sinkkonen and E. Sirén

Electron Physics Laboratory, Helsinki University of Technology, Faculty of Electrical Engineering, Otakaari 7A, SF-02150 Espoo, Finland

Received May 5, 1994; accepted June 15, 1994

Abstract

Two different fabrication techniques to obtain nanometer scale structures without the use of lithography are demonstrated. Quantum dots are made on GaAs by growing strained InP islands by metal-organic vapour phase epitaxy. Quantum confinement of carriers is achieved by the growth of quantum wells on the InP islands. Molecular beam epitaxy is used for the fabrication of a gold island mask on GaAs. Reactive ion etching through the gold mask produces a high density of GaAs columns with diameters down to 20 nm.

1. Introduction

Nanometer scale semiconductor structures, such as quantum wells, quantum wires and quantum dots are currently being investigated because of their novel optical and electronic properties, which make possible new types of semiconductor devices. The fabrication of nanostructures typically involves epitaxy to achieve quantum confinement in the vertical direction. If confinement in more than one dimension is required, the lateral details have to be defined with high resolution lithography, followed by etching or epitaxial regrowth. However, if very high densities of nanostructures are required on large surfaces, the use of lithography becomes very time consuming and difficult. Also, it is not easy to achieve high quality interfaces on etched substrates using epitaxial regrowth techniques. Therefore, it may be advantageous to develop fabrication methods which do not require lithography at all. In this paper we report two such techniques: reactive ion etching (RIE) of nanoscale columns through a random array of gold islands and the maskless, self-organising growth of quantum dots by strained epitaxy using metal-organic vapour phase epitaxy (MOVPE).

One way of fabricating quantum wires is to use dry etching through a suitable mask to produce columns. If the etching is done on a substrate containing a quantum well heterostructure, quantum boxes are obtained inside the column. The mask can be formed without lithography by introducing small particles of suitable size on the surface of the wafer. The use of both colloid particles [1] and aerosol particles [2] has been demonstrated for this purpose. The possible problems with these techniques include the low resistance against etching and a low total density of etched

columns. We have developed a method of forming small gold islands on the GaAs surface. The areal density of the sub-100 nm islands can exceed 10^{10} cm^{-2} and they show good resistance against RIE and are therefore suitable for the fabrication of structures with very high aspect ratios.

The utilisation of lattice mismatch induced 3-dimensional growth is a novel technique of produced structures containing nanoscale islands, providing the means to fabricate quantum box structures with relative ease. Examples of this technique include the formation of InAs or InGaAs dots on GaAs by molecular beam epitaxy (MBE) [3][4], the growth of InAs nanoclusters on InP by MOVPE [5] and the growth of InP islands on GaAs by hydride vapour phase epitaxy (VPE) [6]. We have grown nanoscale InP islands on GaAs by atmospheric pressure MOVPE. Furthermore, selective growth of InGaAs on these InP islands was achieved [7].

2. Reactive ion etching of nanoscale columns using a gold mask

The samples for the gold mask experiments were grown in a VG V80H/V80M MBE system having a deposition chamber for III-V semiconductor compounds connected to a metallisation chamber. Therefore, the structures could be moved after the epitaxial growth from the growth chamber into the metallisation chamber without breaking vacuum. Undoped or n-type GaAs buffers or AlGaAs/GaAs multi-quantum well (MQW) structures, terminated with a thin GaAs cap layer, were grown on either semi-insulating or Si-doped (100) GaAs wafers.

The samples were transferred immediately after the growth into the metallisation chamber. A 1.3 nm thick or 3.0 nm thick Au layer was e-beam evaporated on the fresh GaAs surface at room temperature at a rate of 0.01 nm/s. The thickness of the Au film was measured with an oscillating quartz crystal, calibrated by Rutherford backscattering (RBS) measurement from a thicker Au layer. Because gold grows epitaxially on (100) GaAs, the quality of the surface before and after the Au deposition could be checked with reflection high energy electron diffraction (RHEED). The pattern was streaky after the deposition of Au, showing that the resulting surface is flat. After the gold deposition

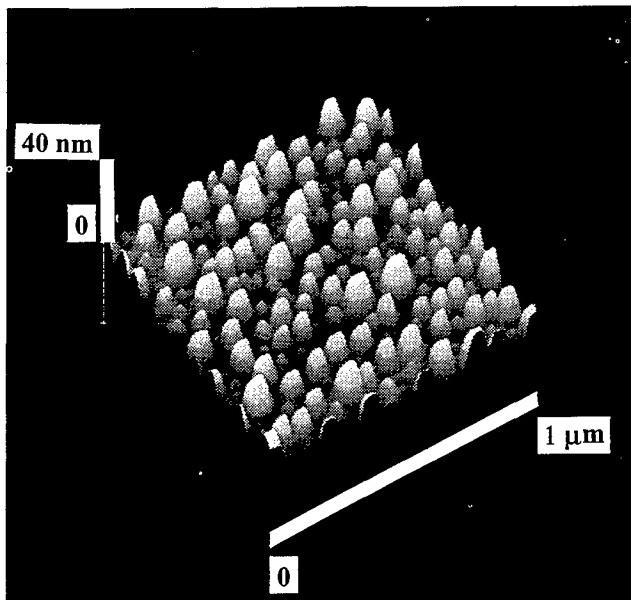


Fig. 1. AFM image of gold islands formed from a 1.3 nm thick gold film on (100) GaAs after a heat treatment at 550 °C.

the samples were heated up to 550 °C in 15 minutes in an arsenic free ambient. The RHEED pattern was observed to change from streaky to spotty after heating the samples above about 250 °C, suggesting a change from a flat to a 3-dimensional surface. At 550 °C the pattern was dim and only some diffuse spots could be distinguished.

The surface of the Au film was analyzed using an atomic force microscope (AFM). Prior to the heat treatment the surface was smooth and featureless. After the heat treatment, however, a high density of small islands is formed on the surface (Fig. 1). The areal density of the islands is $3 \cdot 10^{10} \text{ cm}^{-2}$. The lateral size distribution is quite broad, ranging from 20 nm to 100 nm. Correspondingly, the height of the islands varies from 3 nm to 13 nm.

Reactive ion etching was performed in an Electrotech 320 PC etcher using a mixture of Ar, He and SiCl₄ at a total pressure of 10 mTorr. The power was 40 W and the etching time was varied from 2 to 13 minutes, giving an average etching rate of 1.5 nm/s for GaAs and AlGaAs/GaAs structures. The height of the etched columns was estimated with

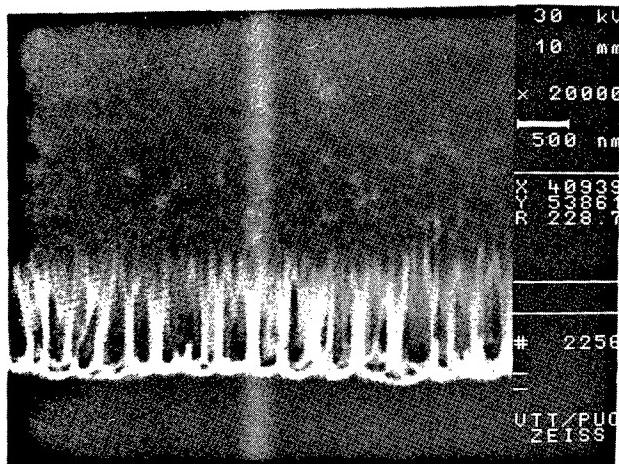


Fig. 2. A SEM micrograph of 1.2 μm high and 60 nm wide AlGaAs/GaAs columns fabricated by reactive ion etching through a gold island mask. The Au mask was formed by *in-situ* heating a 3 nm thick Au film at 550 °C.

a scanning electron microscope (SEM). Shown in Fig. 2 is a SEM micrograph of columns of an AlGaAs/GaAs MQW structure etched through a mask formed from a 3 nm thick Au film. The height of the columns is 1.2 μm and the aspect ratio 20. The shape of the columns shows a slight undercutting in the etching profile. The areal density of the columns is $3 \cdot 10^9 \text{ cm}^{-2}$, which is lower than the original density of the Au islands. This is due to the undercutted etching profile which causes etching through the stem of the smallest columns, homogenising the diameter of the remaining columns. The aspect ratios obtained with this process are as high as 50 and by optimising the etching parameters a further increase in the aspect ratio can be expected.

3. Self-organising growth of InP islands on GaAs by MOVPE

Quantum box samples were grown by atmospheric pressure MOVPE from trimethylgallium (TMGa), trimethylindium (TMIn), tertiarybutylarsine (TBAs) and tertiarybutylphosphine (TBP). The structures were grown on semi-insulating (100) GaAs wafers misoriented by 0°, 2° or 4°. Prior to the growth the substrates were heated up to 700 °C to remove native oxides. First, a 200 nm thick GaAs buffer layer was grown at 650 °C. The InP islands were deposited at 650 °C using TMIn flows of 1.3–3.9 μmol/min and a V/III ratio of 20–50. The deposition time was 4–15 s. The strain caused by the 3.8% lattice mismatch between InP and GaAs induces the growth to occur in a 3-dimensional mode resulting in the formation of small InP islands on the surface [6].

An AFM image of InP islands on a 2° off-oriented (100) GaAs substrate is shown in Fig. 3. The average width of the islands is less than 100 nm and the average height about 20 nm. The areal density of the islands is $3 \cdot 10^9 \text{ cm}^{-2}$. The density of the islands was slightly dependent on the substrate misorientation, being $2 \cdot 10^9 \text{ cm}^{-2}$ for a substrate misoriented by 0° and $3.5 \cdot 10^9 \text{ cm}^{-2}$ for an off-angle of 4°. This dependence of the island density on the misorientation is much weaker in MOVPE than in hydride VPE [8].

Nanoscale InP islands provide a suitable substrate for the selective growth of InGaAs [7]. Therefore, quantum well

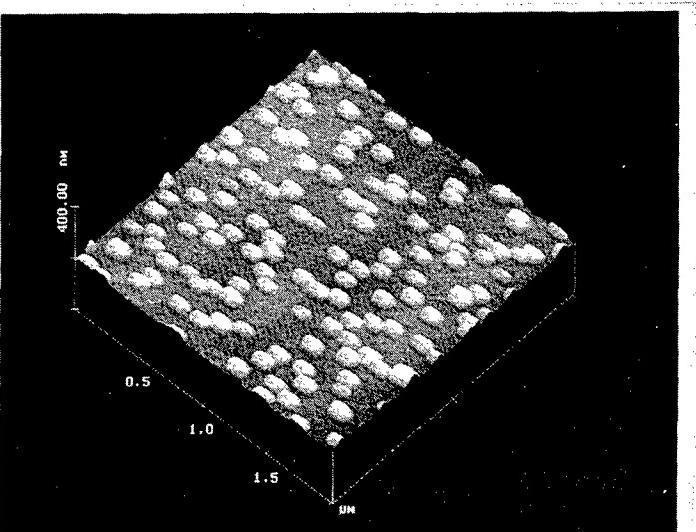


Fig. 3. An AFM image of InP islands grown by MOVPE on a 2° off-oriented (100) GaAs substrate.

structures can be used for improving the carrier confinement in the quantum box. This was verified by high intensity PL emission both at 13 K and at room temperature.

4. Discussion and conclusions

The fact that gold forms islands on GaAs at a relatively low temperature indicates that an eutectic alloy of Au and Ga is formed due to a reaction between the gold film and the substrate. The size of the islands depends on the thickness of the original epitaxial gold film. Our results show that these islands are an excellent mask for the reactive ion etching of nanoscale columns because the selectivity between the Au mask and GaAs/AlGaAs is high. The diameter of the columns is therefore only dependent on the feature size of the mask. The stability of the columns is dependent on the anisotropy, i.e. ion energy, which can be relatively high. An aspect ratio of more than 20 is easily achieved for a column less than 100 nm wide. Moreover, the density of the columns is up to 10^{10} 1/cm² which is the highest value reported anywhere and comparable to the density of electrons in a two dimensional electron gas. The diameter of the etched columns tends to vary due to the random size of the Au islands. If a slight undercutting is accepted in the etching process, the size distribution of the columns can be made more uniform.

The self-organising growth of coherent nanoscale islands on planar substrates is very attractive as a fabrication

method for quantum dot structures. This is because no lithographic or etching steps are necessary. Another advantage is that the structure is grown in a single run without exposing the samples to air, ensuring high quality interfaces. Also, more degrees of freedom for the design of quantum dot structures can be obtained due to the possibility of growing a quantum well selectively on the InP islands. The problems in the method lie in the random distribution of the islands on the surface and in the inhomogeneous size distribution. By optimising the growth parameters the size distribution can be made narrower, and by introducing a thin quantum well on top of the islands the effect of the scatter in the island size can be minimised.

References

1. Iwabuchi, T. *et al.*, SPIE **1284**, 142 (1990).
2. Maximov, I. *et al.*, J. Vac. Sci. Technol. **A11**, 748 (1993).
3. Moison, J. M. *et al.*, Appl. Phys. Lett. **64**, 196 (1994).
4. Leonard, D. *et al.*, Appl. Phys. Lett. **63**, 3203 (1993).
5. Leonelli, R. *et al.*, Phys. Rev. **B48**, 11 135 (1993).
6. Ahopelto, J., Yamaguchi, A. A., Nishi, K., Usui, A. and Sakaki, H., Jpn. J. Appl. Phys. **32**, L32 (1993).
7. Ahopelto, J., Lipsanen, H., Sopanen, M., Koljonen, T. and Niemi, H., Post-deadline paper at the Conference on Indium Phosphide and Related Compounds, Santa Barbara 1994.
8. Ahopelto, J. *et al.*, Extended Abstracts of the 1992 International Conference on Solid State Devices and Materials, Tsukuba, Japan, 1992 (Japanese Society of Applied Physics, Tokyo, 1992) p. 281.

Pulsed Laser Ablation Deposition of CuInSe₂ and CuIn_{1-x}Ga_xSe₂ Thin Films

J. Levoska, S. Leppävuori, F. Wang and O. Kusmartseva

Microelectronics and Material Physics Laboratories, University of Oulu, PL 400, FIN-90571 Oulu, Finland

and

A. E. Hill, E. Ahmed, R. D. Tomlinson and R. D. Pilkington

University of Salford, Dept. of Electronic and Electrical Engineering, Salford M5 4WT, UK

Received May 26, 1994; accepted June 15, 1994

Abstract

CuInSe₂ (CIS) and CuIn_{0.75}Ga_{0.25}Se₂ (CIGS) thin films were deposited by an *in situ* process on fused silica and silicon (100) and (111) substrates by pulsed laser ablation using an XeCl excimer laser and polycrystalline targets. The effect of deposition temperature and substrate material on the structure and orientation of the films was investigated. The structure of the films was studied by X-ray diffraction, and their chemical composition was investigated by energy dispersive X-ray analysis on a scanning electron microscope. The composition of the target material was largely maintained in the films deposited below 450 °C. On fused silica and Si (111) substrates, the films were highly oriented with the (112) planes of the chalcopyrite structure along the substrate surface. The films deposited on to Si (100) developed, under some conditions, {100} orientation and epitaxy.

1. Introduction

CuInSe₂ (CIS) and CuIn_{1-x}Ga_xSe₂ (CIGS) are well established as exceptionally efficient semiconductors with potential applications in the fields of solar cells, infra-red radiation monitors and fibre optic infra-red detectors [1]. The latest versions of the CuInSe₂/Cd(Zn)S heterojunction and the CuInSe₂-based/amorphous silicon tandem solar cell are some of the most efficient thin film photovoltaic devices to have been produced to date. Recently reported thin film CIS cells have achieved efficiencies in excess of 14.8% and module efficiencies are now reported to be 11.7% [2, 3]. Data obtained from these materials indicates a significantly higher degree of radiation resistance and absorption than is found in comparable devices based on either Si or GaAs [4, 5, 6].

Partial substitution of indium by gallium provides a convenient means of adjusting the band gap of the material. As x in CuIn_xGa_{1-x}Se₂ is reduced from 1 to 0 the band gap increases approximately linearly [7] from 1.0 eV to 1.68 eV. Applications for IR detectors would benefit from the small bandgap of CIS with no gallium. This can provide a detector with a response into the 1.3 μm near-IR region of the spectrum, which is of critical importance for fibre optic communications. Solar cell applications require a higher bandgap in order to increase the output voltage and to give an optimum match with the solar spectrum. The composition CuIn_{0.75}Ga_{0.25}Se₂, having a band gap of about 1.17 eV, presents a useful compromise between the optimum band gap of 1.5 eV and ease of film preparation.

Thin films of CIS have been successfully prepared by co-deposition of the constituent elements, either by physical

deposition (PVD) [2] or by chemical vapour deposition (CVD) [8] routes. However, complete control of the process which is necessary in order to achieve reproducible, stoichiometric films is proving difficult to achieve. A simpler approach widely adopted by industry involves the deposition of Cu and In films, typically by electroplating, followed by selenization, and has given good results. However, the latter stage can introduce considerable disturbance to the structure and composition of the films and is also an inherently toxic process with associated environmental implications. For the preparation of CIGS both elemental co-deposition and binary deposition with subsequent selenization have an extra complication in that the introduction of partial gallium substitution of the indium content poses considerable problems of control.

These considerations suggest an alternative route for film deposition whereby a source or target of the required composition acts as a single precursor for the film. Dissociation of the components of such a source has precluded its use in simple PVD evaporation equipment and has forced commercial production to adopt the binary route.

A suitable PVD solution is flash evaporation but preparation of consistent fine grain powder source material presents some problems. Without this control, the production of uniform, pin-hole free thin films with suitable characteristics has proved to be very difficult, although some success with this technique has recently been reported [9].

With the above mentioned deposition processes, there have been problems in the control of homogeneity, composition and carrier concentration in the films [1]. In this study, CIS and CIGS films were prepared by pulsed laser ablation deposition, (PLD) which has proved to be a suitable method for the deposition of multicomponent films, especially in the reproduction of the target composition. PLD has also been used for the deposition of other types of compound semiconductor materials utilised in solar cell technology [10, 11]. Some of our preliminary studies on CIS related films have been published elsewhere [12].

2. Film deposition

The films were deposited by an *in situ* process at temperatures between 50 and 550 °C using an XeCl excimer laser ($\lambda = 308$ nm, 30–50 mJ/pulse on the target). A focused laser beam was scanned over a polycrystalline target held at

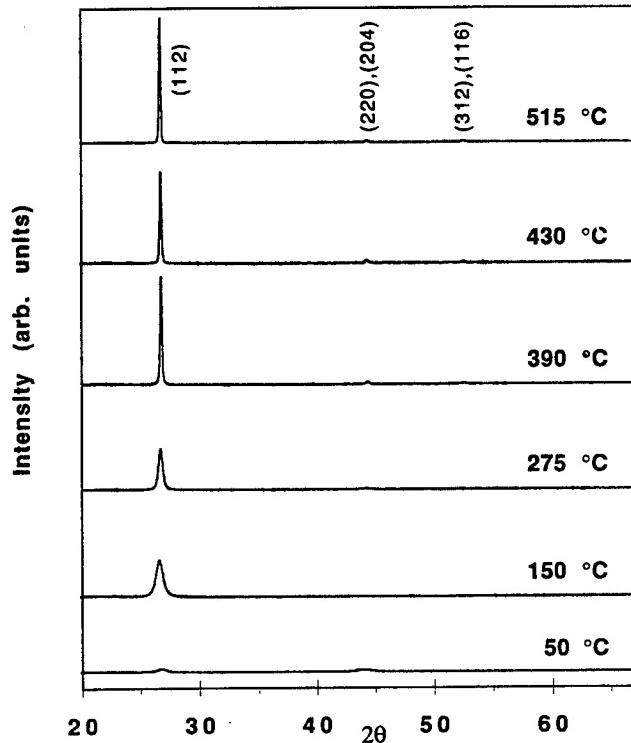


Fig. 1. X-ray diffraction patterns (CuK α) of CIS films deposited at different temperatures on fused silica.

an angle of 45 degrees to the laser beam. The films were deposited on silicon (oxide removed by a HF treatment) and fused silica substrates attached onto a ceramic heater plate by silver paste. The target-to-substrate distance was 3 cm. The ablation process was carried out in a vacuum chamber with a base pressure of about 10^{-5} mbar. The laser beam was focused by a lens outside the chamber, and the fluence level was regulated by defocusing the beam. CIS films were

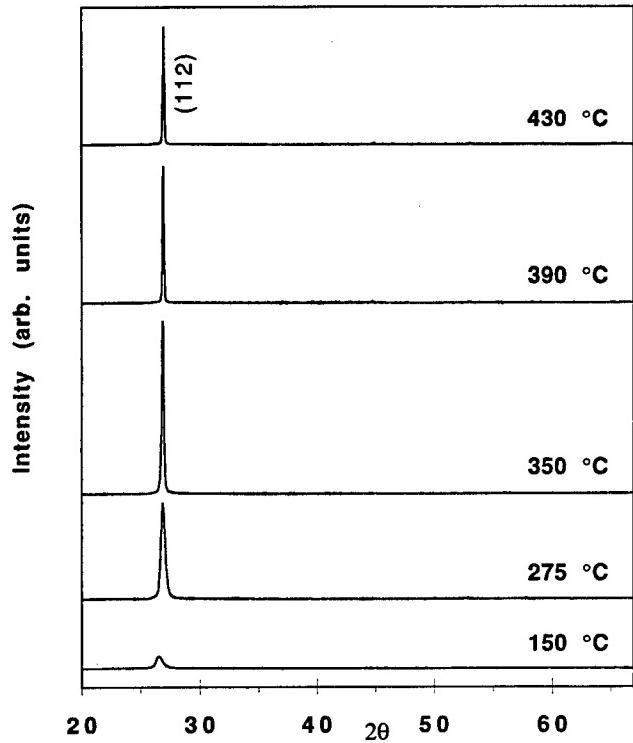


Fig. 3. X-ray diffraction patterns (CuK α) of CIGS films deposited at different temperatures on fused silica.

deposited in a small chamber using a $f = 100$ mm lens (30 mJ/pulse, $4-5$ J/cm 2 , repetition rate 25 Hz). CIGS ($\text{CuIn}_{0.75}\text{Ga}_{0.25}\text{Se}_2$) films were deposited in a modified larger chamber which required the use of a $f = 300$ mm lens and system limitations dictated deposition conditions (50 mJ/pulse, 2 J/cm 2 , repetition rate 10 Hz) which differed from those used for CIS, so the conditions were not directly comparable. After the depositions, the temperature was reduced to room temperature without any additional heat treatments. Typical deposition rates were about 2 Å/s for CIS and 10 Å/s for CIGS and the thickness of the films were 200 ... 450 nm.

3. X-ray diffraction studies

X-ray diffraction patterns ($\theta-2\theta$) were measured using CuK α radiation. For CIS films, Figs 1 and 2, only peaks of the chalcopyrite structure of CIS can be seen. The patterns show that the CIS films deposited even at the lowest temperatures were crystalline, but the X-ray diffraction peaks were broad, indicating low crystal perfection and small crystallite size. The broadness and low intensity of most of the peaks did not allow the chalcopyrite and sphalerite structures to be resolved. The films deposited below 300 °C were oriented with (112) planes of the chalcopyrite structure (or (111) planes of the sphalerite structure) along the substrate surface, independent of the substrate material and orientation. Above 300 °C, the diffraction peaks became narrower and their intensity increased indicating higher perfection, definite chalcopyrite structure and a high degree of preferred orientation. In the films on fused silica, the (112) orientation still remained at the highest substrate temperatures.

Above about 300 °C, the films on Si (100) started to develop {100} type epitaxial orientation which became stronger with increasing temperature. At high deposition temperatures the peaks of the chalcopyrite structure dis-

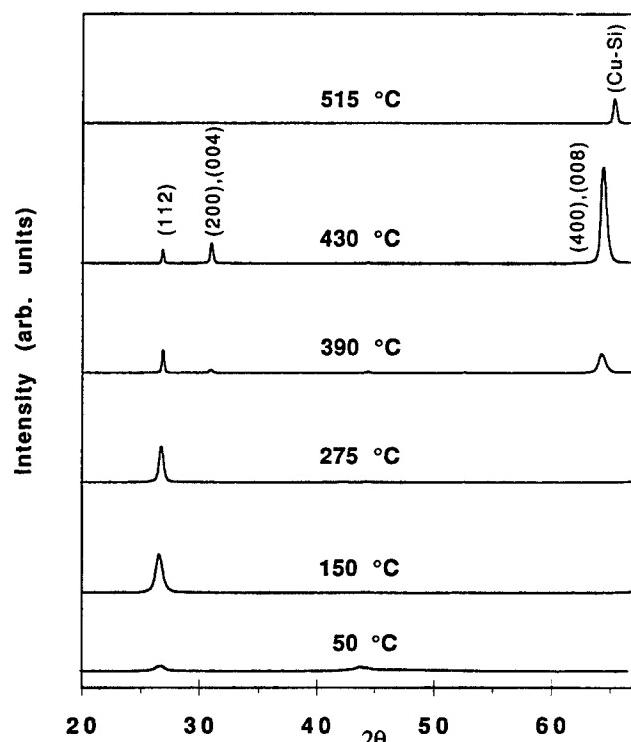


Fig. 2. X-ray diffraction patterns (CuK α) of CIS films deposited at different temperatures on Si (100). (Cu-Si) indicates a phase containing only copper and silicon.

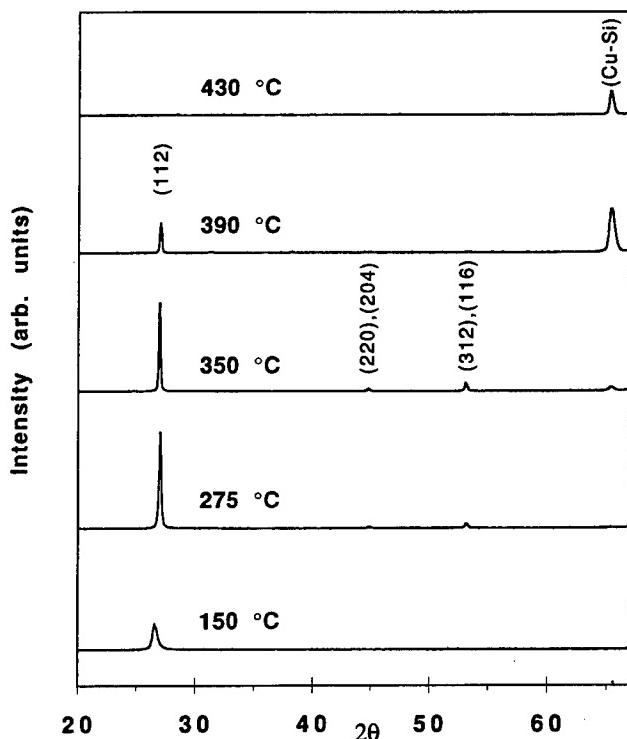


Fig. 4. X-ray diffraction patterns ($\text{CuK}\alpha$) of CIGS films deposited at different temperatures on Si (100). (Cu-Si) indicates a phase containing only copper and silicon.

peared, leaving a discontinuous film containing only Cu and Si. This occurred in the temperature range $450 \dots 600^\circ\text{C}$, depending on the deposition conditions. Preliminary studies showed, that in the films deposited on Si (111), the preferred orientation resembled that on fused silica.

The in-plane epitaxy of the {100} oriented CIS on Si (100) was confirmed by a ϕ -scan through (112) poles, Fig. 5, which shows that the $\langle 100 \rangle$ directions of CIS are oriented along the $\langle 100 \rangle$ directions of the silicon substrate.

The X-ray diffraction patterns of CIGS films deposited at different temperatures on fused silica (Fig. 3) followed the

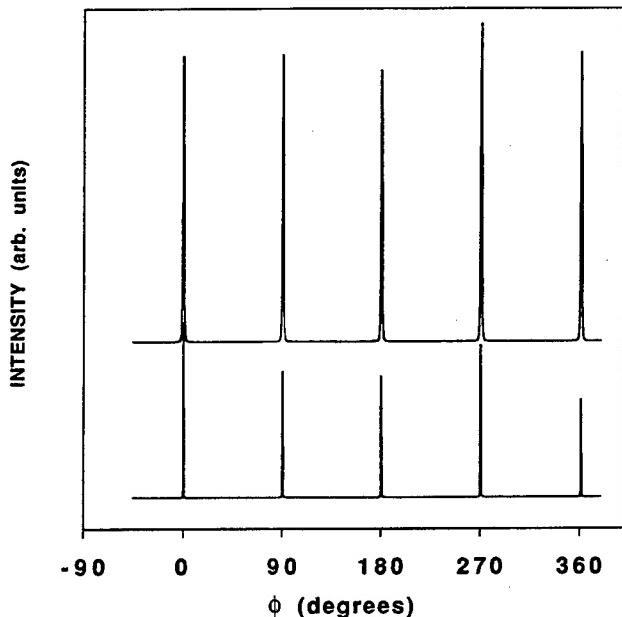


Fig. 5. X-ray diffraction ϕ -scan through (112) poles of a CIS film deposited on Si (100) substrate (upper curve) and ϕ -scan through (111) poles of the Si (100) substrate (lower curve).

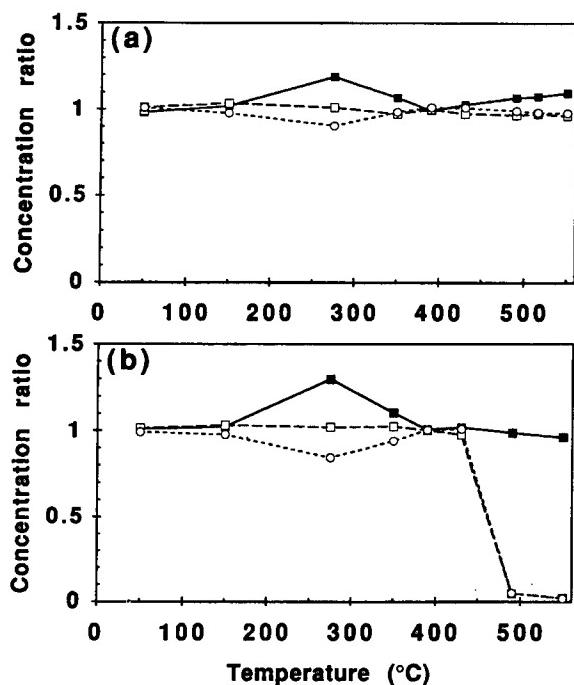


Fig. 6. Chemical compositions of CIS films deposited at different temperatures (a) on fused silica, (b) on Si (100), measured by EDAX. The concentration of each element is presented as a ratio of the concentration in the film/in the target material. Cu: full squares, full line, In: open squares, dashed line, Se: open circles, dotted line.

same general scheme as the CIS films. The diffraction peaks appear at slightly higher angles due to the smaller unit cell of CIGS. In the films deposited on Si (100) (Fig. 4), no development of {100} type preferred orientation was observed at intermediate temperatures (about 400°C for CIS). The disappearance of the (112) oriented chalcopyrite structure and the formation of the phase containing only Cu and Si occurred at lower temperatures than in CIS films. From these results, it is not possible to conclude whether these

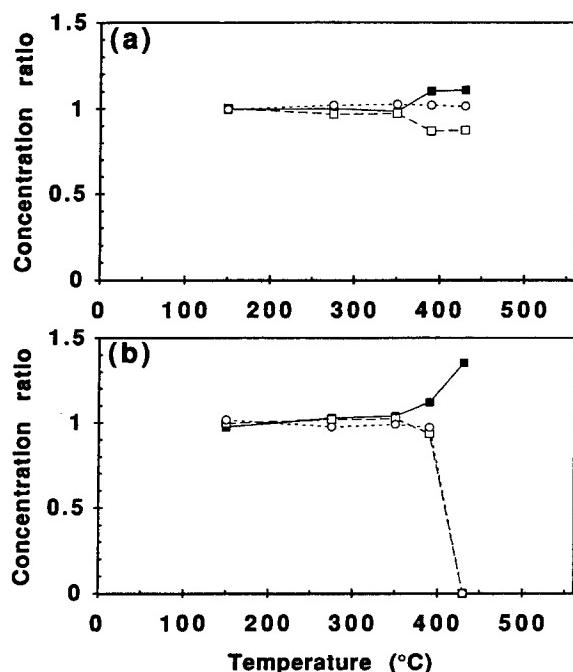


Fig. 7. Chemical compositions of CIGS film deposited at different temperatures (a) on fused silica, (b) on Si (100), as in Fig. 6. Cu: full squares, full line, (In + Ga): open squares, dashed line, Se: open circles, dotted line.

differences were due to chemical composition or different deposition conditions.

The degree of the (112) orientation in CIGS films deposited on fused silica at temperatures between 350 and 400 °C was very strong. For example, in the film deposited at 350 °C, the ratio between the (112) peak and (220, 204) intensities was over 1000, which was higher than in the best CIS films (about 600). These values exceeded the value 70 reported for coevaporated CIS films [13] used in the preparation of high-efficient solar cells [2, 13], and also those reported for sputtered CIS and CIGS films [14].

4. Chemical composition of the films

The chemical composition of the films was determined by energy dispersive X-ray spectroscopic analysis (EDAX) on a scanning electron microscope. The composition and film thickness were calculated by the STRATA thin film program which utilises measurements performed at several accelerating voltages. The elemental concentrations, relative to those of the target materials, are presented in Figs 6 and 7 for CIS and CIGS films, respectively. The composition of the target material was largely maintained in the films deposited on fused silica. Especially, no significant loss of selenium was observed. The Cu/In and Cu/(In + Ga) ratios increased slightly at high temperatures. The reason for the large Cu/Se ratio in the CIS films near 300 °C is not clear. On Si (100) substrates, loss of In and Se (and Ga) occurred above 430 °C in CIS films and at slightly lower temperatures in CIGS films, which causes the disappearance of the chalcopyrite structure, as observed by X-ray diffraction.

5. Conclusions

Films with a good crystal quality, high degree of preferred orientation and good reproduction of the target composition were deposited by pulsed laser ablation. At tem-

peratures below 300 °C, (112) oriented films with a defected chalcopyrite or sphalerite structure were obtained; no effect of the substrate on the structure of the films was found. At higher temperatures the crystal quality improved and the degree of preferred orientation increased, and the substrate material and orientation had an effect on the structure of the films. The constituents of CIS and CIGS reacted with oxide-free Si in the deposition process above 400 °C, so that In, Ga and Se were lost. When considering the reproduction of the target composition and high degree of preferred orientation, deposition temperatures in the range of 350 ... 400 °C seem to be most favorable in the process investigated.

References

1. Rockett, A. and Birkmire, R. W., *J. Appl. Phys.* **70** (1991) R81.
2. Stolt, L. *et al.*, *Appl. Phys. Lett.* **62**, 597 (1993).
3. Ermer, J., Gay, C. R., Pier, D. and Tarrant, D. J., *J. Vac. Sci. Technol. A11*, 1888 (1993).
4. Mickelsen, R. A. *et al.*, *Proc. 18th IEEE Photovoltaics Specialists Conf.* Las Vegas, NV, USA, 1985, p. 1069.
5. Landis, G. A. and Hepp, A. F., *Proc. European Space Power Conf.* Florence, Italy, 1991, p. 517.
6. Tomlinson, R. D. *et al.*, *Journal of Electronic Materials* **20**, 659 (1991).
7. Zegadi, A., Slifkin, M. A., Djamin, M., Hill, A. E. and Tomlinson, R. D., *Physica Status Solidi (a)* **133**, 533 (1992).
8. Jones, P. A., Jackson, A. D., Lickiss, P. D., Tomlinson, R. D. and Pilkington, R. D., *Thin Solid Films* **26**, 4 (1994).
9. Ahmed, E., Hill, A. E., Pilkington, R. D. and Tomlinson, R. D., *J. Phys. D* **26**, 1787 (1994).
10. Feng, Z. C., Perkowitz, S. and Dubowski, J. J., *J. Appl. Phys.* **69**, 7782 (1992), and references therein.
11. Chuu, D. S., Dai, C. M., Hsieh, W. F. and Tsai, C. T., *J. Appl. Phys.* **69**, 8402 (1991).
12. Hill, A. E. *et al.*, *Mat. Res. Soc. Symp. Proc.* **285**, 483 (1993); Levoska, J. *et al.*, *Jpn. J. Appl. Phys.* **32**, Suppl. 32-3, 43 (1993).
13. Bodegård, M. *et al.*, *Proc. 15th Nordic Semiconductor Meeting*, Hämeenlinna, Finland, 1992, p. 281.
14. Yamaguchi, T., Matsufusa, J. and Yoshida, A., *J. Appl. Phys.* **72**, 5657 (1992).

A Study of Dual Conductance Response to Carbon Monoxide of CdS and α -SnWO₄ Thin Films

J. L. Solis,* V. Golovanov,† V. Lantto and S. Leppävuori

Microelectronics Laboratory, University of Oulu, Linnanmaa, FIN-90570, Oulu, Finland

Received May 9, 1994; accepted July 5, 1994

Abstract

An unusual dual conductance response was found in the case of exposure of n-type semiconductors CdS and α -SnWO₄ in the form of thin films to carbon monoxide. Usually, CO behaves as a reducing gas increasing the conductance of n-type semiconductors. However, at temperatures below a border temperature T^* , CO did behave as an oxidizing gas decreasing the conductance of CdS and α -SnWO₄ thin films. T^* for this dual conductance response to CO was found to be below 50 °C for CdS films. The enrichment of the surface with cadmium species had a decreasing effect on T^* which was also decreasing with decreasing partial pressure of CO. For α -SnWO₄ films, T^* was near 250 °C depending on the partial pressure of CO. It was found that the dual response behaviour depends on temperature, CO concentration in air and tin content in the film.

1. Introduction

The chalcogenides CdS and α -SnWO₄ are n-type semiconductors with band gaps typical for insulators. The semiconducting behaviour arises in both structures from native point defects; donors being related to oxygen vacancies in α -SnWO₄ and to cadmium interstitials or to sulphur vacancies in CdS, which is a typical II–VI compound semiconductor. Both SnO₂ and WO₃ are well-known materials for semiconductor gas sensors and have also found applications in commercial sensor constructions. A synthesis of these materials in the form of a new crystal structure of SnWO₄ was first reported in 1972 [1]. The low temperature (α) form of stannous tungstate is stable below 670 °C. In its orthorhombic crystal structure, both metal atoms have distorted octahedral oxygen coordination like in SnO₂ and WO₃. However, differing from the case in SnO₂, tin appears in a divalent form Sn²⁺ in the α -SnWO₄ structure. On the other hand, the wurtzite structure of CdS is based on tetrahedral coordination for both atom types. The direct band gap of CdS is in the visible region (2.5 eV) making the material useful for many optical applications. The optical band gap of α -SnWO₄ was found to be in u.v. region. Illumination only with an u.v. light did increase the conductivity of films, but the peak absorption in the blue (origin of the dark-red colour) had no effect on conductivity.

A survey of the gas response behaviour of a relatively large group of semiconducting oxides is given by Moseley *et al.* [2]. The survey goes further, in that it shows that materials and gas responses can be classified to reveal generalisations comprehensible in terms of proposed models for

gas response. They also describe [2, 3] an unusual dual conductance behaviour for some semiconducting oxides (they call it transition between n-type and p-type behaviour). However, stannous tungstate is missing from the survey and we have not found any data of gas response properties of α -SnWO₄ in the literature. Here we describe the dual conductance response of CdS and α -SnWO₄ thin films to carbon monoxide which usually behaves as a reducing gas, increasing the conductance of n-type semiconductors. For comparison, measurements were also made with a pure WO₃ film and with some Sn_xWO_y films having lower tin concentrations ($x < 1$).

2. Experimental

An electro-hydro-dynamical technique was used for the deposition of CdS thin films from CdCl₂ and (NH₂)₂CS solutions on a heated (480 °C) glass substrate [4]. A variation of the stoichiometric composition of CdS films was obtained by mixing the initial solutions in different proportions of Cd/S ions. XPS (X-ray Photoelectron Spectroscopy) was used for the determination of the atomic Cd/S ratios at the film surfaces. Three different film types with the Cd/S surface ratios of 3.2 (I-type), 2.4 (II-type) and 1.6 (III-type) were used in the study. The average grain size of the films was found to be around 100 nm from AFM (Atomic Force Microscope) images. Evaporated indium was used for making ohmic contacts with the films.

Stannous tungstate films were grown by means of reactive co-sputtering with a Balzers BAS 450 magnetron sputtering system, where a tin target was in d.c. mode and tungsten target in r.f. mode, respectively. Both thermally oxidized Si(111) and glass were used as substrates for the films. The deposition sputtering was made in an argon atmosphere containing 11% of oxygen. Different sputtering powers (75–150 W) were selected to the tin target and 150 W to the tungsten target in order to obtain films with different Sn/W ratios. The samples used in the study are identified in Table I with the sputtering parameters used in deposition and with the Sn/W ratio (x in Sn_xWO_y) and the thickness d of the films. Three different temperatures (room temperature, 100 °C and 200 °C) were chosen for the substrate during deposition. The thickness and composition of the films were obtained from RBS (Rutherford Backscattering Spectroscopy) and EDS (Energy Dispersive Spectroscopy of X-rays) results, respectively. The films were annealed at 400 °C or at 600 °C for four hours in air.

XRD (X-ray Diffraction), CEMS (Conversion Electron Mössbauer Spectroscopy) and Raman Spectroscopy

* Permanent address: Facultad de Ciencias, Universidad Nacional de Ingeniería, P.O. Box 1301-Lima, Peru.

† Permanent address: Faculty of Physics, Odessa State University, Pastera 42, 270100 Odessa, Ukraine.

Table I. Identification of different Sn_xWO_y thin films ($x = \text{Sn}/W$ ratio) with sputtering parameters (powers to W and Sn targets and substrate temperatures) used in deposition and with the Sn/W ratio and thickness d of the films

Sample	W/r.f. (W)	Sn/d.c. (W)	T (°C)	Sn/W ratio	d (nm)
SnW1	150	150	29	1.12	135
SnW2	150	150	200	0.65	170
SnW3	150	130	29	0.48	170
SnW4	150	100	100	0.37	150
SnW5	150	75	29	0.27	130
SnW6*	150	100	29	0.03	110
W1	150	—	29	0.00	135

* Deposition with mask.

together with SEM and AFM imaging were used for characterization of stannous tungstate films. The average grain size was found to be around 200 nm from AFM images. It was found that as-grown films after deposition were amorphous, but an annealing at 400 °C did crystallize the films. The films, deposited on substrates at room temperature with a sputtering power of 150 W to both targets (SnW1 in Table I) were found to contain polycrystalline α -SnWO₄ in an orthorhombic structure as the major phase and a small amount of SnO₂. The sample SnW6 with a very small amount of Sn was very similar to pure WO₃.

A decomposition of α -SnWO₄ into SnO₂ and WO₃ phases after annealing at 600 °C in air was found from both XRD and CEMS results. XRD and CEMS together made possible a very satisfactory identification of phases in the films. In addition, Raman spectra were also measured from the films up to 1900 cm⁻¹ and results were compared with the measured Raman spectra from SnO₂ and WO₃ films. Raman spectra from CdS films were typical of the wurtzite structure.

Two-point method was used in electrical conductivity measurement and wire contacts to stannous tungstate films were made with a low-temperature gold paste. A computer-controlled measuring system employing the flow-through principle was used for varying gas concentrations, heater voltages of the films and for data acquisition, handling and storage [5]. A d.c. voltage of 1 V was used in the conductance measurements of the films.

3. Results

3.1. CdS films

Figure 1 shows the conductance response Gg/Go above the border temperature T^* and Go/Gg below T^* to 100 ppm of CO in dry nitrogen (oxygen and humidity concentrations few ppm) for the three different CdS film types I, II and III. Go is the measured conductance in nitrogen and Gg the conductance after exposure to CO. CO behaves as a reducing gas at all temperatures with the III-type film, but has border temperatures T^* at 300 K and 320 K with the I- and II-type films, respectively. With these two film types the behaviour of CO changes from reducing to oxidizing when the temperature of the films decreases through T^* . The enrichment of the surface by Cd (I-type films) had a decreasing effect on T^* . An increase of the sensitivity to CO with increasing surface content of Cd is also shown in Fig. 1, both above and below T^* .

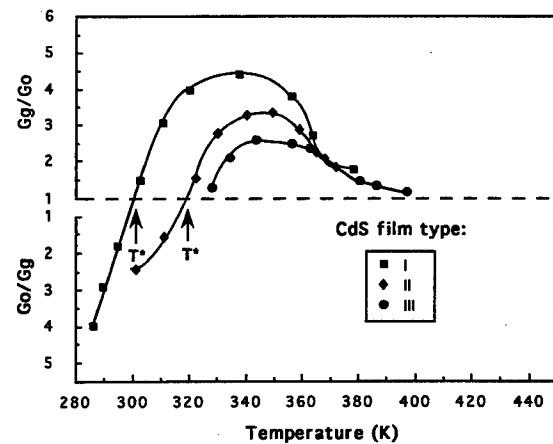


Fig. 1. Conductance response Gg/Go (and Go/Gg) of I-, II- and III-type CdS films to 100 ppm of CO in nitrogen at different temperatures. T^* is the border temperature between oxidizing and reducing behaviour of CO.

Figure 2 shows typical response (τ_{res}) and recovery (τ_{rec}) times (corresponding to a change of 90% of the total change) of I-type films in the case of exposure to 100 ppm of CO in nitrogen at different temperatures. The temperature T^* has an important role in the temperature behaviour of both τ_{res} and τ_{rec} . Especially, τ_{res} has a high jump at T^* .

IR spectroscopy has been used in order to study possible changes in the chemisorption of CO on CdS surfaces above and below the temperature T^* . Typical results in the absorption region of carbonilic species, 2000–2300 cm⁻¹, are shown in Fig. 3 for I-type and III-type films. An exposure of I-type films to 100 ppm of CO in nitrogen at 290 K led to an immediate appearance of a broad and intense absorption peak in the region 2085–2115 cm⁻¹. The peak consisted of two components at 2090 cm⁻¹ and 2110 cm⁻¹. At the same time, a decrease was observed in the conductivity of films. After removing of CO from the atmosphere, the band at 2110 cm⁻¹ disappeared first, while the band at 2090 cm⁻¹ decreased gradually and vanished after 15 min. A shift of the absorption peak up to 2140 cm⁻¹ was found after increasing the film temperature above T^* to 340 K. IR spectra, measured from III-type films, did show only a less intense peak at 2140 cm⁻¹ [Fig. 3(b)].

3.2. Sn_xWO_y films

Figure 4 shows the temperature dependence of conductance, in the form of an Arrhenius plot, measured for different

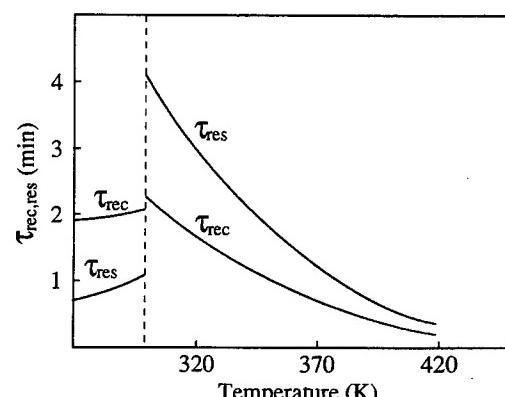


Fig. 2. Response (τ_{res}) and recovery (τ_{rec}) times of an I-type CdS film at different temperatures for 90% conductance changes after introducing and removing 100 ppm of CO in nitrogen, respectively.

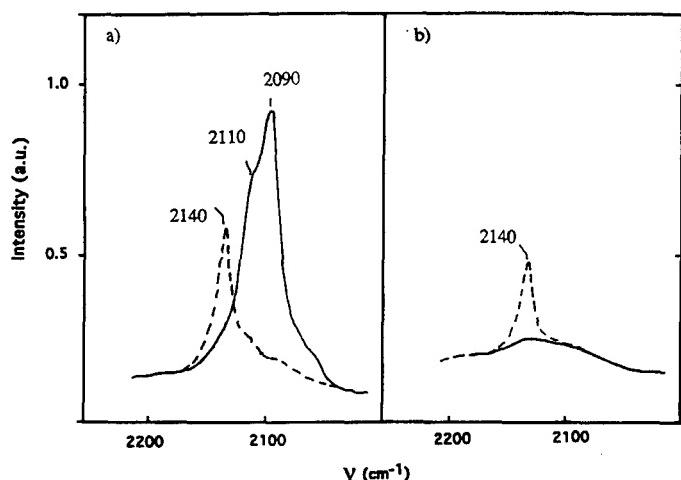


Fig. 3. IR absorption bands in the region of carbonilic species, measured after CO chemisorption on surfaces of (a) I-type and (b) III-type CdS films at 290 K (solid line) and 340 K (dashed line).

Sn_xWO_y films in Table I at a heating rate of 2.4 °C/min in synthetic air and in nitrogen (few ppm of oxygen and humidity) between room temperature and 400 °C. An increase of Sn content in the films did increase the conductivity of the film, especially in synthetic air at low temperatures near room temperature. Simultaneously, the temperature coefficient of conductivity decreased. The increase of conductance of the α - SnWO_4 film (SnW1) in synthetic air in Fig. 4(a) corresponds to an average activation energy of 0.18 eV up to 400 °C. Large differences appear in conductance values of the pure WO_3 film (and SnW6 film)

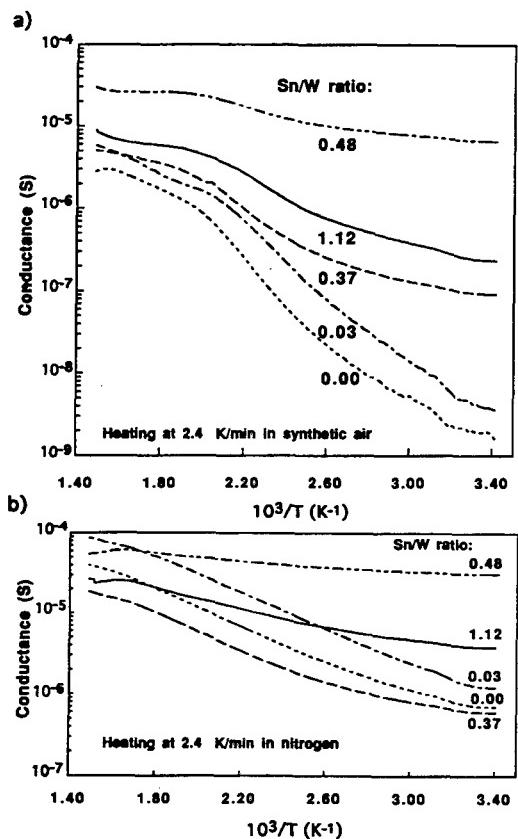


Fig. 4. Conductance of different Sn_xWO_y films in Table I ($x = \text{Sn}/\text{W}$ ratio), measured (a) in dry synthetic air and (b) in nitrogen at different temperatures between room temperature and 400 °C.

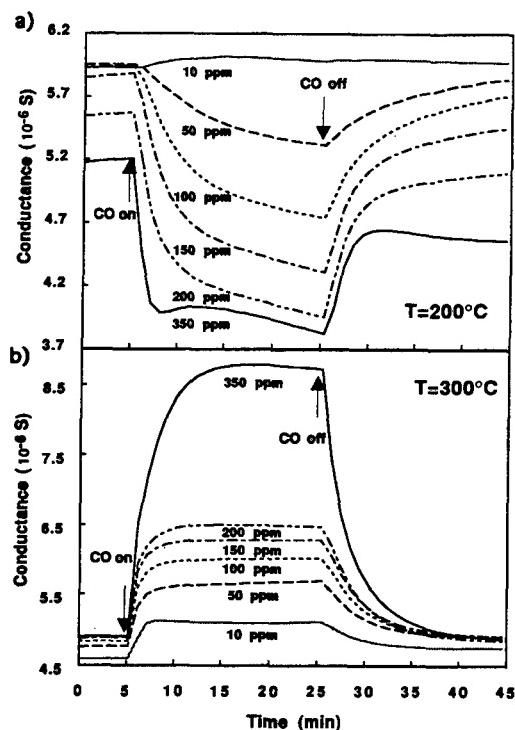


Fig. 5. Conductance response of an SnW1 (α - SnWO_4) film to different concentrations of CO in dry synthetic air (a) at 200 °C and (b) at 300 °C.

between synthetic air and nitrogen at low temperatures in Fig. 4.

Figure 5 shows the conductance response of an SnW1 (α - SnWO_4) film to different concentrations of CO in synthetic air at 200 °C and 300 °C. A dual behaviour of conductance response to CO is shown between 200 °C and 300 °C in Fig. 5. At 300 °C, conductance of the film increased with increasing CO concentrations, but at 200 °C, conductance increased at exposure to 10 ppm of CO while, at higher CO concentrations, conductance decreased with increasing CO concentration. A dual conductance behaviour was now observed also at a constant temperature of 200 °C by varying CO concentration. As is shown in Fig. 5, both response and recovery times have large differences between 200 °C and 300 °C.

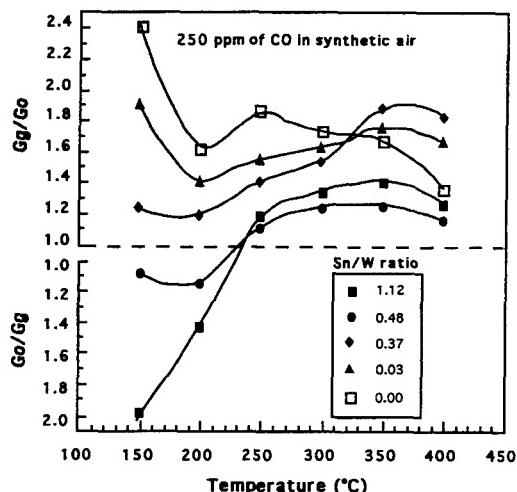


Fig. 6. Conductance response G_g/G_o (and G_o/G_g) of different Sn_xWO_y films in Table I ($x = \text{Sn}/\text{W}$ ratio) to 250 ppm of CO in synthetic air at different temperatures.

It was found that dual conductance response to CO does not depend only on temperature and CO concentration, but also on the tin content in Sn_xWO_y films. This is shown in Fig. 6 where the conductance ratios between synthetic air (Go) and 250 ppm of CO in synthetic air (Gg) are plotted for five different films in Table I, having different tin concentrations, in a temperature range between 150°C and 400°C. Only the two films SnW1 and SnW3 show dual conductance behaviour with respect to temperature. The border temperature T^* is about the same, around 230°C, to both films. α - $SnWO_4$ was found from SEM and AFM images to be the matrix phase in these films. A strong increase of the conductance ratio Gg/Go of the pure WO_3 film (and SnW6 film) at 150°C is shown in Fig. 6.

4. Discussion

The role of the adsorption/desorption mechanism and that of the surface defect mechanism in creation of the transducing conductance signal in semiconductor gas sensors is, to some extent, in a controversial state. The creation of oxygen surface vacancies in oxidic semiconductors is of central importance in the surface defect mechanism. In the case of conductance response to CO or to changes in oxygen partial pressure, oxide materials like α - $SnWO_4$ may reflect changes in the amount of oxygen vacancies (donors) in addition to changes in the amount of adsorbed oxygen, while the first changes are not possible in the case of CdS films.

The oxidizing behaviour of CO, in connection with the dual conductance response, is very difficult to explain. Its appearance in both oxidic α - $SnWO_4$ and CdS films may rule out changes in the amount of donors as the origin for the behaviour. Different electrode materials used for CdS (indium) and α - $SnWO_4$ (gold) films may also rule out electrodes as the origin for the behaviour [6]. Then, the origin should be at the film surfaces. The abrupt changes in the response and recovery times (Fig. 2) and in CO chemisorption (Fig. 3) at the border temperature T^* also support this reasoning. A possibility is that CO behaves as a surface acceptor trapping electrons to its antibonding $2\pi^*$ orbital. The red shift in the carbonyl stretching band in Fig. 3(a), when the CO behaviour changes from reducing to oxidizing below T^* , is in agreement with that interpretation. Changes in polar surface layers of ionic semiconductors and corresponding changes in the electron affinity of the semicon-

ductor may serve as another explanation for the dual conductance behaviour in response to CO.

The results in Fig. 1 refer to CO response of CdS films in nitrogen (few ppm of oxygen), while the results in Fig. 5 refer to CO response of an α - $SnWO_4$ film in synthetic air. It seems that the dual conductance behaviour is not sensitive to the amount of oxygen in ambient atmosphere. In fact, a dual conductance behaviour was also found by CdS films in the case of response to CO in synthetic air. A large difference of about 200°C is shown between the temperatures T^* of CdS and Sn_xWO_y films in Figs 1 and 6, respectively. This is a typical temperature difference also for the similar sigmoid behaviour of conductance versus temperature [Fig. 4(a)] between CdS and oxidic semiconductor films. With increasing temperature, above room temperature, the increase rate of conductance of CdS and oxidic semiconductor films starts to decrease at around 50°C and 250°C, respectively. An unusual conductance response of some SnO_2 thick films to CO after a fast cooling process is described in Ref. [7]. The situation with those SnO_2 films resembles the present situation where the unusual dual conductance behaviour appears only by films with high cation concentrations (Fig. 1) and high conductivity at temperatures below T^* (Fig. 4).

Acknowledgements

The authors wish to thank Mr. J. Frantti and Mr. J. Remes for their advice in EDS, SEM and AFM experiments. J. Solis expresses high gratitude to International Science Programs at the Uppsala University, Sweden for the financial support, and V. Golovanov expresses his gratitude to the Centre for International Mobility, Finland for the financial support.

References

1. Jeitschko, W. and Sleight A. W., *Acta Cryst.* **B28**, 3174 (1972).
2. Moseley, P. T., Stoneham, A. M. and Williams, D. E., in: "Techniques and Mechanisms in Gas Sensing" (Edited by P. T. Moseley, J. O. W. Norris and D. E. Williams) (Adam Hilger, Bristol/New York 1991), pp. 108–138.
3. Moseley, P. T., *Sensors and Actuators* **B3**, 167 (1991).
4. Golovanov, V., Serdiouk, V., Stys, L., Thchemeresiouk, G. and Shmiltovich, A., *Ukrain. Phys. J.* **33**, 157 (1988) (in Russian).
5. Romppainen, P. and Lantto, V., Report S93 (Department of Electrical Engineering, University of Oulu, Oulu, Finland 1987), 22 pp.
6. Ylinampa, A., Lantto, V. and Leppävuori, S., *Sensors and Actuators* **B13–14**, 602 (1993).
7. Lantto, V. and Rantala, T. S., *Sensors and Actuators* **B5**, 103 (1991).

A Cluster Approach for Modelling of Surface Characteristics of Stannic Oxide

Tuomo S. Rantala¹, Vilho Lantto¹ and Tapiro T. Rantala²

¹ Microelectronics and Material Physics Laboratories, University of Oulu, Linnanmaa, FIN-90570 Oulu, Finland

² Department of Physics, University of Oulu, Linnanmaa, FIN-90570 Oulu, Finland

Received May 9, 1994; accepted in revised form July 5, 1994

Abstract

An ab initio cluster approach (LDA-SCF) is used to study the electronic structure of the SnO_2 (110)-1 \times 1 surface without any relaxation and reconstruction. Modelling is made for the ideal nonpolar surface and also for surfaces containing different kinds of oxygen vacancies, foreign impurities and adsorbates. Computations are based on the linear combination of atomic orbitals (LCAO), which allows the population analysis through projection of the molecular orbitals onto the atomic basis set. This is used to trace the atomic character of the cluster levels. A cluster with 103 atoms $\text{Sn}_{32}\text{O}_{71}$ is used to model the ideal SnO_2 (110) surface, the most stable and thus the dominant crystallite face of the polycrystalline SnO_2 , and the next few bulk layers. The results here refer to both the ideal nonpolar (110) surface and some oxygen-deficient polar (110) surfaces. Chromium and NO are taken as examples of impurity atoms and adsorbate molecules, respectively. Based on the results we are able to suggest an explanation for the observed strong decrease of conductivity of SnO_2 polycrystalline films and the enhanced adsorption and dissociation of NO on the SnO_2 surface due to adding of chromium impurities.

1. Introduction

Stannic oxide, SnO_2 , in its pure form is an n-type wide band-gap semiconductor. Its electrical conductivity results from point defects which are native (mainly oxygen vacancies) or foreign atoms that act as donors or acceptors. Some unique electrical and optical properties of SnO_2 make it useful for many applications, like gas reduction and detection. For instance, the majority of semiconductor gas sensors today use SnO_2 as the reactive element.

Local energy levels, originating from defects like oxygen vacancies at or near SnO_2 surfaces, may have an important role in both electrical transport and gas response properties of SnO_2 [1]. Oxygen vacancies behave as double donors in the bulk SnO_2 . However, the theoretical prediction from band-structure calculations [2] that oxygen vacancies at the surface of SnO_2 do not act as donors is supported by some UPS and conductance measurements [1]. Similar results were also obtained from our earlier cluster model [3], used for the simulation of the surface effects on some energy levels of SnO_2 in relation to oxygen vacancies.

Chromium impurities in SnO_2 have a strong effect on both its electrical conductivity and catalytic properties. Incorporation of small amounts (1% or less) or Cr into the surface layer of SnO_2 basically influences the surface properties of SnO_2 and results in a very effective catalyst for the low-temperature reduction of NO [4]. It seems likely that the high reactivity of Cr ions at the surface is responsible for the enhanced adsorption and dissociation of

NO. Only 0.1 wt% of Cr_2O_3 powder mixed with the SnO_2 powder has been shown to cause, after sintering, the conductance of some SnO_2 thick-film samples to drop by several orders of magnitude [5]. Due to the very high resistance values, the samples were not applicable for conductance measurements below 500 °C.

A cluster model of up to a hundred or more atoms will be shown to be able to simulate the characteristics of a large crystal surface, also reflecting the energy levels of the surface with band formation. The cluster model can also include foreign atoms at the solid surface or surface defects like oxygen vacancies in oxidic semiconductors. Furthermore, the model gives a detailed picture of the energy and shape of the individual orbitals of the surface atoms and their modification by the presence of impurities and adsorbates.

2. Computational method

In general, localised structures and phenomena of bulk solids or solid surfaces can be conveniently modelled with finite clusters. Point defects, foreign impurities and adsorbates on the surfaces are typical examples of such objects, as they induce local charge redistributions affecting on the atomic bonding network and electronic structure. Cluster models, on the other hand, do not reproduce the three dimensional band structure of the infinite and periodic solid, and therefore, special expertise may be needed in the model building and interpretation. Furthermore, the finite size effects should be under careful control and the following questions should be answered. What are the effects from the cluster boundaries, or conversely, what would be the effects of embedding the cluster onto the solid surface or bulk? What are the effects from the charge unbalance or lack of screening, if ionic species or strongly polarized clusters are considered? Should we always use stoichiometric clusters? The relevance of these questions vary within materials from metals to semiconductors and insulators and could, of course, be often simulated with some kind of embedding process [6].

One of the important advantages of the cluster model is the easy calling of self-consistent (SCF) ab initio methods [7]. Also semiempirical methods could be employed allowing the treatment of larger clusters but possibly being more limited in the description of the electronic structure. We have chosen to use the linear combination of atomic orbitals (LCAO) method and local-density approximation [8]. This makes it possible to obtain a detailed picture of the bond formation and modifications, all based on the description

using atomic orbitals. LCAO allows us to project the electronic levels of the cluster onto the atomic orbitals, which are used as basis functions. This is the most transparent way of explaining bond formation or breaking in adsorption or vacancy formation in terms of atomic orbitals. Therefore, the minimal basis set tin orbitals up to $5p$ but without $4f$ and oxygen orbitals up to $2p$, was used. On the other hand, the core orbitals do not participate in the bonding. The tin $4d$ and $5s$ orbitals are 27 and 10 eV and oxygen $1s$ and $2s$ orbitals are 510 and 24 eV below the vacuum level, respectively. Hence, the frozen core approximation was adopted for tin orbitals in the K, L, M and N shells and for oxygen orbitals in the K shell. The justification of the frozen core approximation for these levels was verified using smaller clusters for comparative studies.

Ab initio methods (including electronic correlations) are computationally heavy. However, employment of the symmetry properties in the cluster arrangement reduces much of the labour, and the local-density approximation (LDA) of the density-functional theory makes the computational approach of relatively large clusters attractive. For very large systems or infinite solids the LDA is the only practical method which can be used to include the exchange and correlation effects. For computations we have used the DMol code [8] which is now also commercially available [9] and von Barth-Hedin exchange-correlation functional [10], which is simple and straightforward to use also in the spin-polarized form.

3. Cluster model

Figure 1 shows the unit cell of the rutile crystal structure of SnO_2 containing two Sn and four O atoms. The structure is 6 : 3 coordinated and has a relatively strong ionic character. The bulk electronic-structure calculations for SnO_2 [11] predicted a direct optical band gap of 3.6 eV in agreement with experimental results. The conduction-band minimum, which is about 4.5 eV below the vacuum level, is 90% Sn s-like state and surprisingly similar to that for the free-electron model. The valence-band maximum region of width 2 eV consists mainly of oxygen p lone-pair orbitals directed perpendicularly to the Sn-O bonds. In some calculations the optical band gap has been found to be nearly free of surface states [12] and surface defect states originating from different oxygen-vacancy models at the (110) surface [2]. However, the experimental results from UPS measurements [13] show that oxygen-deficient SnO_2 (110) surfaces display

a significant density of surface states in the band gap up to the Fermi level, but oxidized (more ideal) surfaces show no significant density of band-gap states.

Our computations refer to the SnO_2 (110) face which is the most stable, and thus the dominant crystallite face of SnO_2 . The use of the (110) surface without any reconstructions in our computations is based on experimental findings of the thermally driven reconstructions of ion-bombarded SnO_2 (110) surfaces in Ref. [14]. The (110) crystal plane is shown in Fig. 1 and Fig. 2 illustrates the cluster $\text{Sn}_{32}\text{O}_{71}$ used for modelling the atomic structure of the ideal (110)- 1×1 surface. The cluster $\text{Sn}_{32}\text{O}_{71}$ in Fig. 2 with a C_{2v} symmetry has 103 atoms and in addition to that two atoms of a diatomic adsorbate over a surface tin (site 1). In the [110] direction, the crystal is seen to be composed of nonpolar "unit" layers, each containing three atomic planes, $[(\text{O}^{2-})(2\text{Sn}^{4+} + 2\text{O}^{2-})(\text{O}^{2-})]$ per (110) unit area. On ideal surfaces, containing also the "bridging" oxygens (98 and 100 in Fig. 2), there are equal numbers of fivefold- and sixfold-coordinated surface cations. The "bridging" oxygens appear as rows in the [001] direction at the outermost plane and occupy bridging positions between second-plane sixfold-coordinated tin atoms. Oxygen atoms in the second plane of the outermost nonpolar "unit layer" (20 and 24 in Fig. 2) are referred to as "in-plane" oxygens. The top view of the ideal (110) surface in the [110] direction is shown in Fig. 3.

Here the results refer only to different modifications of the basic cluster in Fig. 2, although smaller clusters were also considered, especially for adsorbate studies. Some molecular computations were carried out at the higher level, too. The

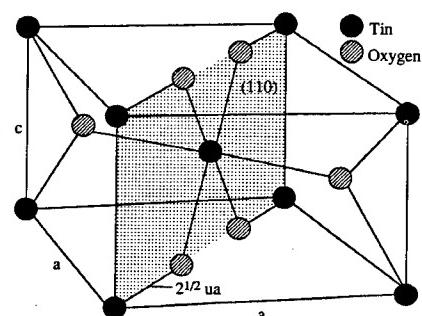


Fig. 1. Unit cell of the rutile structure of SnO_2 with the tetragonal D_{4h}^{14} symmetry and parameter values $a = 4.737 \text{ \AA}$, $c = 3.186 \text{ \AA}$ and the internal parameter $u = 0.306$ (each cation has two anions at the distance $\sqrt{2}ua$ in the surrounding distorted octahedron).

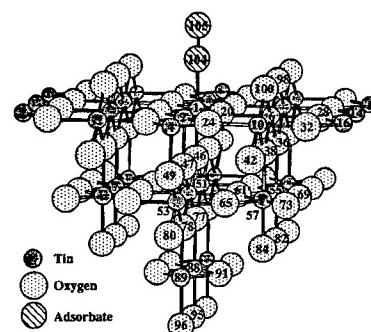


Fig. 2. Structure of the basic cluster $\text{Sn}_{32}\text{O}_{71}$ with C_{2v} symmetry and with a diatomic "adsorbate" (sites 104 and 105). This is the model for the ideal SnO_2 (110)- 1×1 surface.

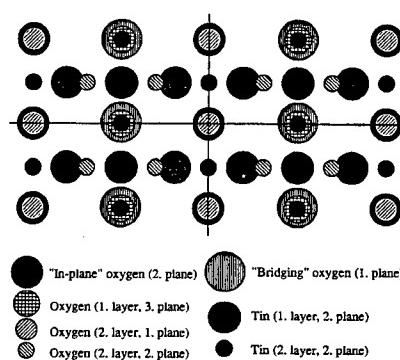


Fig. 3. Top view of the ideal SnO_2 (110)- 1×1 surface in the [110] direction illustrating the positions of tin and oxygen atoms in the five uppermost surface planes in two "unit" surface layers.

basis set was extended to include more functions to reveal any notable basis set effects. The spin-polarised treatment of molecular orbitals was allowed in search for the ground state of the molecule and to determine its contribution to the molecular energetics.

For the computational reasons the cluster size was limited to about one hundred atoms. A typical SCF run for such a cluster with 115 iteration cycles on Convex C3840 took about 500 minutes. Therefore, only molecular conformations and adsorbate positions on the surfaces were optimized by minimizing the total energy, but any reconstruction of the cluster geometry was not allowed. Experimental results in Refs [13, 14] show that this is a very realistic assumption for the SnO_2 (110) surface. Any relaxation of the boundary layer atoms at the cluster surface could have enhanced the finite size effects only, instead of leading to the true local reconstructions.

4. Results and discussion

The 5s level of the free tin atom is about 10 eV below the vacuum level, but the oxygen octahedron around a tin atom in bulk SnO_2 pushes the bottom of the bulk conduction band of SnO_2 (90% Sn s-like state) up to about 4.5 eV below the vacuum level. Here we concentrate mainly on the energy and the origin of the cluster levels around the "band gap" region. Computed energy levels from the basic cluster in Fig. 2 (without any adsorbate) are shown in Fig. 4 at this energy region (zero energy in all Figs 4, 5 and 6 refers to the vacuum level). The symbols of the energy levels indicate the strongest atomic-orbital component from the projection together with the atom label from Fig. 2 as a subscript. For some mixed levels, also the second strongest component is given above the strongest. For example, the symbol $5s_{10}$ means that 5s orbital of the tin atom 10, but also the 5s orbitals of three other tin atoms because of the C_{2v} symmetry. The energy level with two full circles (spin degeneration) is the highest fully occupied level, while the level with two open circles is the lowest fully unoccupied level in the computation. The levels between these two have

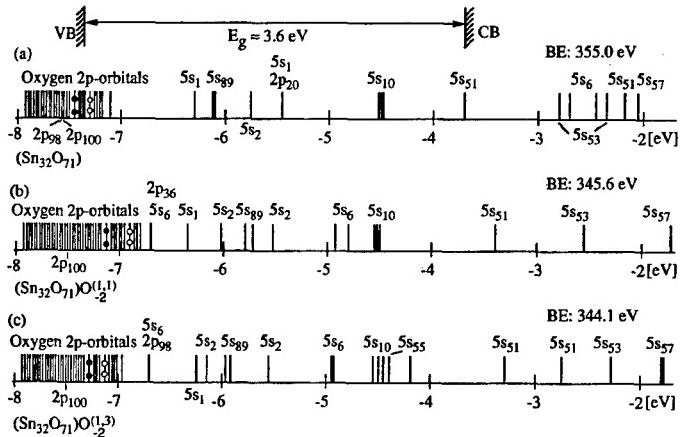


Fig. 4. Computed energy level diagrams around the "band-gap" region ($E_g \approx 3.6$ eV) (a) of the basic cluster $\text{Sn}_{32}\text{O}_{71}$ in Fig. 2 and of its two modifications. The modifications contain the basic cluster with two oxygen vacancies at sites 98 and its C_{2v} symmetry site (b) and at sites 36 and its C_{2v} symmetry site (c). The energy values of the levels are given with reference to the vacuum level. The cohesion energies of the clusters (binding energies, BE) are given with reference to the free atoms.

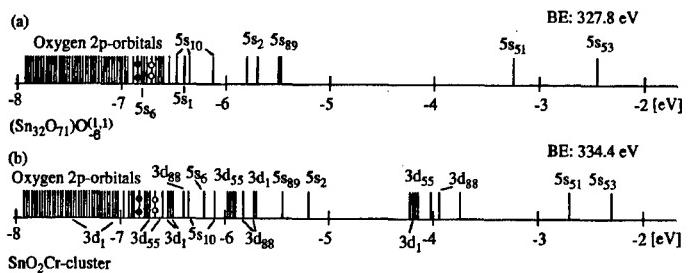


Fig. 5. Computed energy level diagrams around the "band-gap" region of (a) a cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}^{(1,1)}_6$ where all six "bridging" oxygens (sites 98, 100 ...) in the basic cluster are removed, and of (b) its modification SnO_2Cr -cluster where four tin atoms at sites 1, 55 and its C_{2v} symmetry site and 88 are replaced by chromium impurities. The vacuum level and the free atoms are the references for the energy levels and binding energies (BE), respectively.

a partial population as an indication of the thermal distribution of occupations. The total cohesion energies with reference to those of free atoms (binding energies, BE) are also given in the figures. The conduction-band minimum is connected to the $5s_{51}$ (5s orbital of tin atom 51) level (-3.7 eV), since Sn at the site 51 may be the best representative for the bulk tin in the cluster. The valence-band maximum in Fig. 4 is fixed to the top region of oxygen $2p$ orbitals corresponding quite well to the true band-gap value of 3.6 eV. Some surface states, originating from Sn and O atoms at the sites 10 and 20 at the (110) surface, are also shown in the "band gap".

The origin of the local energy levels, related to oxygen vacancies in SnO_2 , is in the neighbouring tin atoms and, especially, in their 5s orbitals. Electronic structures of two clusters, both with two oxygen vacancies in the surface layer, are also shown in Fig. 4. The cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}^{(1,1)}_2$ has two oxygen vacancies in the 1. plane of the 1. layer (the "bridging" oxygens 98 and its C_{2v} symmetry partner in Fig. 2 are missing). In the cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}^{(1,3)}_2$ the two oxygen vacancies are in the 3. plane of the 1. layer, when the two oxygen atoms 36 and its C_{2v} symmetry partner in Fig. 2 have jumped to the "bridging" vacancies in the cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}^{(1,1)}_2$. Sn at the site 6 is the neighbouring tin atom to the oxygen vacancy in both clusters. New energy levels in the "band gap", originating from 5s₆ orbitals, are

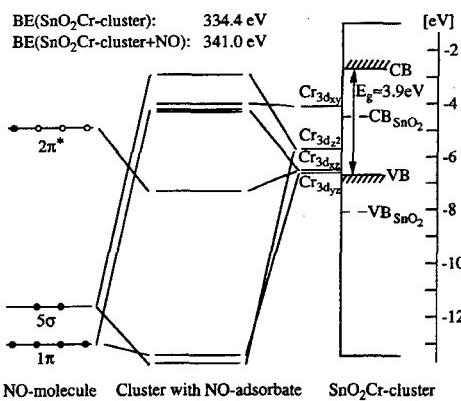


Fig. 6. Computed energy level diagrams of the NO molecule, the cluster SnO_2Cr with four Cr impurities [Fig. 5(b)] and of the interaction of NO molecule with the Cr impurity on the SnO_2 (110) surface (site 1 in the cluster). The vacuum level and the free atoms are the references for the energy levels and binding energies (BE), respectively. Band edges CB and VB for both the cluster and bulk SnO_2 (---) are also shown in the figure. The subscripts of Cr 3d orbitals describe different symmetries of these orbitals.

shown in Fig. 4. The levels are at the bottom of the "band gap" and also above the middle point of the "band gap" in agreement with experimental findings in Ref. [13]. There are also some differences in the "electron affinity" between the three clusters in Fig. 4. After removal of a "bridging" oxygen, for instance, its extra electrons may reside at the tin atom below it (Sn^{2+} instead of Sn^{4+}) making some polarity to the "unit" surface layer.

All six "bridging" oxygens in the cluster in Fig. 2 are removed in the cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}_{-6}^{(1,1)}$. The energy levels of this cluster are shown in Fig. 5 together with levels for its modification where four tin atoms (sites 1, 55 and its C_{2v} symmetry partner, and 88 in Fig. 2) are replaced by chromium impurity atoms. The energy levels originating from 5s orbitals of the tin atoms below the removed "bridging" oxygens in the cluster $(\text{Sn}_{32}\text{O}_{71})\text{O}_{-6}^{(1,1)}$ are near the "valence-band" edge, which means that these tin atoms at the oxygen-deficient (110) surface are in the divalent form Sn^{2+} . The band-gap levels, observed to continue up to the Fermi level in the case of oxygen-deficient (110) surfaces [13], may arise from "in-plane" oxygen surface vacancies. The "band-gap" levels, originating from 3d orbitals of Cr impurities in Fig. 5, are in two groups. One is near the bottom of the "gap" and the other above the middle point of the "gap". The levels originating from the surface chromium ($3d_1$) are in the lower group. Deep trap levels in the band gap, originating from Cr bulk impurities, decrease the concentration of conduction electrons in n-type SnO_2 . The levels from the Cr surface impurities, near the bottom of the band gap, trap electrons and increase the surface energy barrier of this n-type semiconductor. Thus, both bulk and surface impurities may have a very strong effect on conductivity of polycrystalline films.

Figure 6 shows energy level diagrams describing the interaction of the NO molecule with the Cr impurity on SnO_2 (110) surface, computed from the cluster SnO_2Cr with four Cr impurities in Fig. 5(b) together with the NO molecule above the surface chromium (site 1 in Fig. 2). Three most important energy levels 1π , 5σ and $2\pi^*$, computed without spin polarization for NO, are also shown in Fig. 6. A smaller cluster was used to model the SnO_2 (110) surface with a Cr impurity when the position of the NO molecule above the surface impurity was optimized. This position was used in the computation of the electronic structure in Fig. 6 with the large cluster. The energy difference between the bonding 1π and antibonding $2\pi^*$ orbitals has decreased from the value 8.1 eV in the free NO molecule to 6.1 eV as a result of NO adsorption. The transient formation and dissociation of NO^- species has been assumed to be important in

the catalytic reduction of NO [4]. The transfer of an extra electron from surface Cr to the antibonding $2\pi^*$ orbital of NO leads to the significant weakening of the N-O bond. The results in Fig. 6 are in good agreement with the above reasoning. It may be possible to relate the deep surface traps (mixed $2\pi^*$ states in Fig. 6) with the high reactivity of Cr surface impurities to explain the enhanced adsorption and dissociation of NO.

5. Conclusions

Band-gap levels were found to arise from "bridging"-oxygen vacancies at the SnO_2 (110) surface in agreement with experimental findings in Ref. [13]. The levels were near the valence-band edge when "bridging" oxygens were removed from the (110) surface (Fig. 5). Chromium bulk impurities were found to create deep trap levels in the band gap which may decrease concentration of conduction electrons in n-type SnO_2 . The levels from chromium surface impurities were near the bottom of the band gap. These levels can trap electrons and increase the surface band bending of SnO_2 . Thus, both types of Cr impurities are likely to decrease conductivity of SnO_2 polycrystalline films. An electron transfer from chromium surface impurity to adsorbed NO molecule may weaken the N-O bond and explain the enhanced adsorption and dissociation of NO on SnO_2 surfaces having chromium impurities.

References

1. Kohl, D., in: "Gas Sensors" (Edited by G. Sberveglieri) (Kluwer, Dordrecht 1992), pp. 43–88.
2. Munnix, S. and Schmeits, M., J. Vac. Sci. Technol. **A5**, 910 (1987).
3. Rantala, T. S., Lantto, V. and Rantala, T. T., Sensors and Actuators **B19**, 716 (1994).
4. Solymosi, F. and Kiss, J., J. Catal. **41**, 202 (1976).
5. Sberveglieri, G. et al., Sensors and Actuators **B1**, 79 (1990).
6. Langreth, D. and Suhl, H. (Editors), "Many-Body Phenomena at Surfaces" (Academic Press, Orlando, Florida 1984).
7. Rantala, T. T., Acta Universitatis Ouluensis **A184**, Physica **29** (University of Oulu, Oulu 1987).
8. Rantala, T. T., Wästberg, B. and Rosén, A., Chem. Phys. **109**, 261 (1986).
9. DMol User Guide, Version 2.2 (Biosym Technologies, San Diego, CA 1992).
10. von Barth, U. and Hedin, L., J. Phys. **C5**, 1629 (1972).
11. Robertson, J., J. Phys. **C12**, 4767 (1979).
12. Munnix, S. and Schmeits, M., Phys. Rev. **B27**, 7624 (1983).
13. Cox, D. F., Fryberger, T. B. and Semancik, S., Phys. Rev. **B38**, 2072 (1988).
14. deFrésart, E., Darville, J. and Gilles, J. M., Appl. Surf. Sci. **11/12**, 637 (1982); Solid State Commun. **37**, 13 (1980).

The Ionisation Process of α Particles in Mesoscopic Structures: Simulation by Monte Carlo Method¹

K. Tarnay,^{a*},^b F. Masszi,^b T. Kocsis,^a A. Poppe^a and L. Kiss^c

^a Technical University Budapest, Department Electron Devices, H-1521 Budapest, Hungary

^b Scanner Laboratory, Electronics Department, Institute of Technology, Uppsala University, P.O. Box 534, S-751 21 Uppsala, Sweden

^c Uppsala University, Institute of Technology, Department Solid State Physics, P.O. Box 534, S-751 21 Uppsala, Sweden

Received May 9, 1994, accepted in revised form July 7, 1994

Abstract

A first principle based quasi-deterministic 3D particle dynamics Monte Carlo simulation method was developed for examining mesoscopic (sub-half micron) Si electron devices. Applying a novel method for calculating the field and potential distributions, the real trajectories of the carriers are exactly followed. Consequently, an important feature of this method is that all Coulomb scatterings are inherently taken into account. A description of the physical background, the models and the simulation principle is given. The boundary conditions and a deterministic model for Auger recombination is also presented. Finally a simulation example is described: the evaluation of a dense hole-electron plasma induced by an α particle, passing through a reverse biased pn-junction. In connection with this example we also detail the model applied for the carrier generation induced by α particles penetrating the device.

1. Introduction

The following considerations have led us to apply the particle dynamics Monte Carlo method for simulation of sub-half micron semi-conductor structures:

- Assuming a Si device structure with a volume of $0.25 \times 0.25 \times 0.25 \mu\text{m}^3$ and a doping density of 10^{23} m^{-3} , the number of ionised impurities is only about 10^3 . The number of carriers is of the same order of magnitude.
- The classical methods of semiconductor device simulation (drift-diffusion or hydrodynamic semiconductor equations) — are based on a continuum view and apply certain statistical considerations for the carrier distribution functions². These methods are not valid any more if the number of carriers in the simulated structure, like in our case, lies only in the order of magnitude of a few thousands or less.

The relatively small number of carriers suggests the development of a particular 3D Monte Carlo simulation method for studying the mesoscopic device behaviour, where the trajectories of each carrier are individually and exactly followed both in the real space and in the k -space (wave vector space). The increasing computing power of the up-to-date

computers allows to perform such a simulation by using only a reasonable amount of CPU-time. Considering various Monte Carlo simulation methods, the particle dynamics method seems to be the most suitable technique³. More sophisticated Monte Carlo methods (by applying charge clouds, superparticles etc.) can offer far more effective numerical solution tools, but the physics of the simulated system can easily be obscured.

The simulation in the 3D space can sometimes be simplified by a 2D approach (i.e. a unit-width 3D system, assuming perfect uniformity in the neglected direction). The required computing power could sufficiently be reduced in this way. This simplification cannot be applied for a particle dynamics Monte Carlo simulation of charged particles, because the point charges in 2 dimensions can only be implemented as line charges. In this case we would get an electric field of $1/r$ dependence and a potential of logarithmic dependence, instead of $1/r^2$ and $1/r$ behaviour, respectively. Since the force acting to the charged carriers is proportional to the field, the calculated trajectories in any 2D case would be sufficiently different from the real 3D ones.

Our 3D particle dynamics Monte Carlo semiconductor device simulator is based on the concepts of classical physics (Newton law, Coulomb law etc.). The fact, that the carriers are elementary particles, is taken into consideration only by anisotropic effective masses, and by limitations in the momenta (velocities) by Bragg reflections, to remain in their first Brillouin zone. In this way, the bulk behaviour can be modelled quite accurately. However, in some cases the wave nature of the carriers is not negligible⁴.

During the development of the method we concentrated to apply the possible deepest first physical principles inside the examined structure.

Using the Monte Carlo method, the simulation results are the instantaneous states and scattering events of each individual carrier (like carrier position, carrier velocity, carrier

¹ The research was sponsored by Digital Equipment Co. (External European Research Projects HG001 and SW-003, Swedish Board of Technical Development (NUTEK), Swedish Institute and Hungarian Research Foundation (OTKA).

* e-mail: tarnay@eet.bme.hu.

† e-mail: ferenc@sim.Teknikum.uu.se

² Derived from the Boltzmann transport equation (applying the relaxation time approximation and considering the first few moments of the electron distribution function).

³ In particle dynamics, all particles — i.e. in our case charged carriers, dopant ions, interface charges — are treated as point charges, without performing any charge assignment to elementary volumes.

⁴ The modelling of quantum mechanical effects in a potential well near the Si-SiO₂ interface of a MOS transistors is different, i.e. the wave nature of the carriers will be dominant. In theory, a self-consistent solution of the field and the Schrödinger equations may give correct results, but the computer time requirement in the 3D real and momentum spaces is extremely large — for the up-to-date supercomputers, too. Thus, for this case, some other approach is needed. A detailed description is given in [1].

wave vector, carrier energy, time instant of a scattering event, etc.). The quantities used in classical⁵ semiconductor device theory (such as carrier concentrations, mobility of carriers, current densities or scattering rates etc.) can be derived as time- and space averages of the quantities delivered by the Monte Carlo simulation method.

The development of a particle dynamics Monte Carlo semiconductor device simulator represents two different kinds of tasks:

1. The development of the simulator itself.
2. The development of tools to interpret the results of the Monte Carlo simulation in terms of the classical semiconductor device physics. This gives the possibility to compare the results to experimental data or to the results given by classical semiconductor device simulators. These tools can be used to apply Monte Carlo simulation results to optimal parameter determination for classical semiconductor device simulators.

2. The principle of the method

The state of carriers is characterised by the real space coordinate vector r and by the wave vector k .

The carrier dynamics simulation is based on the effective mass concept. As long as the constant effective mass concept is a good approximation, the results give an exact description of the real physical phenomena.

Figure 1 shows the flow chart of the simulation.

2.1. Dispersion relations.

The dodecaeder-shaped first Brillouin zone of silicon is approximated by a sphere (Debye approximation), having a radius of

$$k_{\text{Max}} = 2\pi/a_{\text{Si}} \quad (2.1)$$

2.1.1. Dispersion relation for electrons. For electrons, there are six ellipsoid-shaped constant energy surfaces in the first Brillouin zone. For $\langle 100 \rangle$ oriented silicon, the principal axes of these ellipsoids are located on the positive or negative coordinate axes in the k -space and their centres are situated at $0.854 k_{\text{Max}}$ [2].

The dispersion relation for electrons in the i -th valley is

$$W_i = \frac{1}{2}\hbar^2(k - k_{0i})\mathbf{m}_{\text{eff}(i)}^{-1}(k - k_{0i}) \quad (2.2)$$

where $\mathbf{m}_{\text{eff}}^{-1}$ is the reciprocal effective mass tensor, which has the following generic form:

$$\mathbf{m}_{\text{eff}}^{-1} = \begin{bmatrix} \frac{1}{m_x} & 0 & 0 \\ 0 & \frac{1}{m_y} & 0 \\ 0 & 0 & \frac{1}{m_z} \end{bmatrix} \quad (2.3)$$

We assume that

- the reciprocal effective mass tensors are diagonal, i.e. all off-diagonal elements are zero.

⁵ In the following the terms "classical semiconductor device theory" or "classical semiconductor device simulators" are used in that sense, that they are based on the Boltzmann kinetic equation (drift-diffusion or hydrodynamic semiconductor equations).

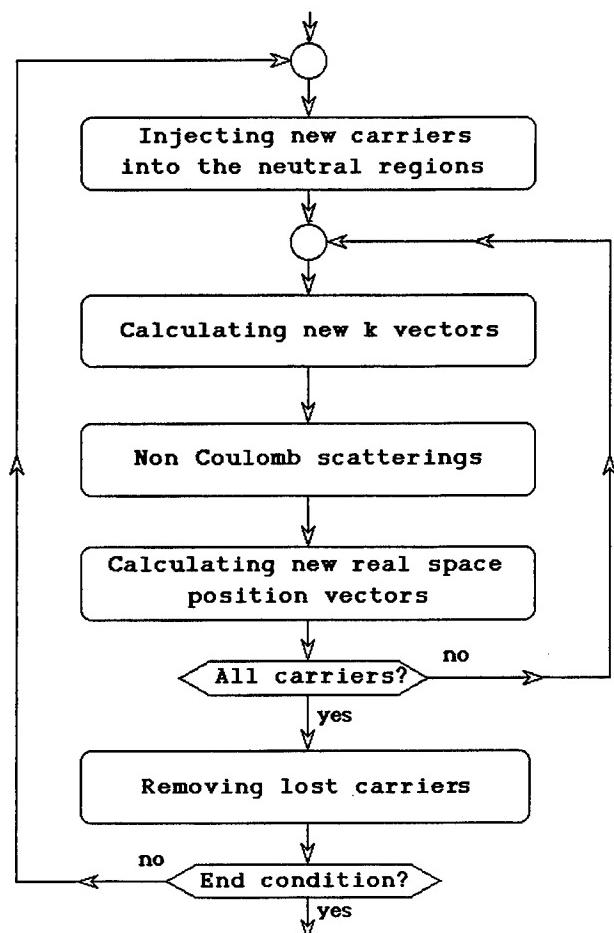


Fig. 1. The flow chart of the simulation.

- the diagonal elements of the reciprocal effective mass tensor are assumed to be independent of the k -vector (and of particle energy) having a value corresponding to the zero energy constant energy surface.
- As usual approximation, the diagonal elements of the effective mass tensor may have two different values, the transverse effective mass m_{\perp} and the longitudinal effective mass m_{\parallel} . For example for the $\langle 100 \rangle$ orientation $m_x = m_{\parallel}$, $m_y = m_{\perp}$, $m_z = m_{\perp}$.

Since the zero energy constant energy surfaces are not situated at the origo of the k -space, the electron momentum and velocity vectors are linearly dependent on their k -vectors. The relationship between the k -vector, velocity and momentum for the electrons of the i th ellipsoid is given by

$$\mathbf{p} = \mathbf{m}_{\text{eff}(i)}^{-1} \mathbf{v} = \hbar(\mathbf{k} - \mathbf{k}_{0i}) \quad (2.4)$$

2.1.2. Dispersion relation for holes. The masses of heavy and light holes correspond to two concentric spherical constant energy surfaces (i.e. the warped shapes of the valence band constant energy surfaces are approximated by spheres). Caused by the assumed spherical symmetry of the constant energy surfaces, the hole effective masses are considered to be isotropic, thus the diagonal elements of the hole effective mass tensor are equal, i.e. a scalar effective mass can be considered. The dispersion relation for holes is given by

$$W = \frac{1}{2}\hbar^2 \frac{|\mathbf{k}|^2}{m_{\text{eff}}} \quad (2.5)$$

Different effective masses are used for the light and heavy holes (m_l and m_h , respectively). The centre of the constant energy is at the origin of the k -space, thus the hole velocity vector is directly proportional to the k -vector.

Split-off holes are not taken into consideration.

2.3. Carrier dynamics

The state of carriers in the real space and momentum space is described by the motion equation. During a short time interval the change of the carrier state is caused

- by forces acting to the carrier, and
- by non-Coulomb scattering events⁶ (see later).

2.3.1. The motion equation. The carrier dynamics is described by the classical Newtonian law of motion

$$\frac{d^2\mathbf{r}}{dt^2} = \mathbf{m}_{\text{eff}}^{-1} \mathbf{F}(\mathbf{r}) \quad (2.6)$$

where $\mathbf{F}(\mathbf{r})$ is the force acting to the carrier.

The forces acting to the carriers can be split into two components:

$$\mathbf{F}(\mathbf{r}) = \mathbf{F}_p(\mathbf{r}) + \mathbf{F}_\infty(\mathbf{r}) \quad (2.7)$$

where

— $\mathbf{F}_p(\mathbf{r})$ is the Coulomb force originating from charges inside the structure i.e. from charged carriers and from fixed charges (donors, acceptors and charged interface states). This force is evaluated by using the Coulomb law

$$\mathbf{F}_p(\mathbf{r}_i) = \frac{1}{4\pi\epsilon} \sum_{j \neq i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|^3} \quad (2.8)$$

resulting in an electric field component

$$\mathbf{E}_p(\mathbf{r}) = \pm \frac{q}{4\pi\epsilon} \sum_j \frac{\mathbf{r} - \mathbf{r}_j}{|\mathbf{r} - \mathbf{r}_j|^3} \quad (2.9)$$

which corresponds to a potential component

$$\psi_p(\mathbf{r}) = \pm \frac{q}{4\pi\epsilon} \sum_j \frac{1}{|\mathbf{r} - \mathbf{r}_j|} \quad (2.10)$$

— $\mathbf{F}_\infty(\mathbf{r})$ is the force caused by the field of external voltages

$$\mathbf{F}_\infty(\mathbf{r}_i) + \mp q_i \mathbf{grad} \psi_\infty \quad (2.11)$$

This component is evaluated by the boundary value problem solution of the Laplace equation

$$\operatorname{div} \mathbf{grad} \psi_\infty = 0 \quad (2.12)$$

(with a set of boundary conditions modified with respect to that of usually applied in the solution of the Poisson equation).

The first integral of the motion equation (2.6) yields the carrier velocity vector, and the second integral results in the carrier position vector. Applying a time increment Δt small enough to assume a constant force during this time interval, the integration for determining the carrier trajectories can be carried out by first order numerical quadrature formulae⁷. For the velocity space:

$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t) + \mathbf{F}(\mathbf{r}, t) \Delta t \quad (2.13)$$

⁶ The Coulomb scattering events are inherently accounted for by the solution of the motion equation (i.e. depending upon the forces).

⁷ Because of the field is a strongly varying function of the position coordinate, using the lowest order formulae with small time step Δt gives better accuracy than higher order formulae with larger time step.

and for the real space:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t) \Delta t + \frac{1}{2} \mathbf{F}(\mathbf{r}, t) \Delta t^2 \quad (2.14)$$

The velocity and momentum space coordinates are connected according to the relation (2.4).

Initial values of the integrations are the carrier position and velocity vectors of the previous simulation step. The advantage of the described method is that all Coulomb scatterings (carrier-ionised impurity, carrier-carrier and carrier-charged interface state) are exactly followed⁸ in a deterministic way.

The effect of forces arising from the magnetic field of moving charges is neglected.

2.3.2. Potential calculation. In classical semiconductor device simulation methods, the potential distribution $\psi(\mathbf{r})$ is usually the primary quantity and is calculated by solving the Poisson equation. The electric field $\mathbf{E}(\mathbf{r})$ is a derived quantity, given as the negative gradient of the potential $\psi(\mathbf{r})$. For obtaining a carrier density, a carrier assignment to grid elements is necessary⁹.

The difficulties for the particle dynamics method with the Poisson equation are the following:

- If a carrier is assigned to a grid element with a given weight function (w), the original charge q and the assigned charge distributions ($\delta\rho$) are different, thus the potential and the electric field acting on the carriers will be different as well. This causes a self-accelerating electric field [3]. Even if we have just a single electron, there will be a given electric field due to the effective charge density assigned to the grid element where the electron resides. Since the electric field accelerates all charged particles, our single electron would accelerate itself.
- Carriers very close to each other are affected by the same electric field and cannot feel each other's attracting or repelling force. If the velocity and the motion direction of these carriers are only slightly different, they can travel together in the structure unless one of them is scattered by some scattering process.
- Another problem arises from the small number and non-uniform distribution of point charges: the assigned charge density can be very rough and oscillating. This leads to numerical instabilities in the solution of the Poisson equation, slowing down to its convergence.

In our approach these problems are inherently avoided, since the ψ_p potential component originating from the point charges is a derived quantity, see eq. (2.10) and the ψ_∞ component describing the effect of boundary conditions is given *a priori*. The sum of these potential components gives the total potential

$$\psi(\mathbf{r}) = \psi_p(\mathbf{r}) + \psi_\infty(\mathbf{r}) \quad (2.15)$$

The potential arising from the charges inside the structure should be evaluated only at the sides of the examined structure (they are needed for the modified boundary conditions of the Laplace equation). The potential inside the structure – if needed – is calculated in the post processing phase of the simulation from charge positions.

⁸ Assuming that the time step Δt is small enough.

⁹ Without charge assignment the required charge density for the Poisson equation is unknown, so the Poisson equation cannot be solved.

The advantage of this method is self-explanatory: without solving Poisson's equation, which would result in self-accelerating carriers, as shown earlier, the exact field (force) caused by other point charges is calculated analytically for all carrier positions. Only the field (force) stemming from the boundary conditions is calculated by the numerical solution of the Laplace equation. Since the spatial distribution of the Laplace potential is smooth enough, this field (force) component can be calculated quite accurately.

2.3.3. Boundary conditions and carrier behaviour at the boundaries. Since in the outlined methods for the potential and field distribution calculations analytical and numerical methods are mixed, some new problems have arisen, not existing in conventional simulators using pure numerical methods. Although the boundary problems can be separated (boundary conditions for the potential and for handling of carriers leaving the structure at the boundaries), these two parts cannot be treated independently because they are inherently connected.

The shape of the examined structure is a rectangular octaeder. Depending upon the structure to be analysed, the treatment of the six boundary sides should be different.

For some boundaries, artificial boundary conditions are used. These appear always for two opposite sides of the structure, and are similar to a periodic boundary condition. For such boundary planes the potential is averaged in the direction perpendicular to the examined plane, and the averaged potential is applied as a Dirichlet boundary condition, for example, for the y -direction

$$\psi(x, 0, z) = \psi(x, W, z) = \frac{\sum_{j=0}^{n_y} \psi(x, y_j, z)}{n_y} \quad (2.16)$$

If a particle leaves the structure at a boundary, it will return in the opposite side, i.e. if the calculated position coordinate is outside the structure at a point

$$P_0(x, y, z) \quad y > W \quad (2.17a)$$

the particle will be placed to a new position of

$$P'_0(x, W - y, z) \quad (2.17b)$$

For pn junction structures (Fig. 2).

(a) artificial boundary conditions are applied for the $x - z$ planes at $y = 0$ and at $y = W$ and for the $y - z$ planes at $x = 0$ and at $x = L$.

(b) Dirichlet boundary conditions are applied for the $x - y$ planes at $x = 0$, assuming a potential of 0 and at $z = D$, assuming a potential of $U_{\text{ext}} + \phi_{\text{diff}}$. The widths of the p and n depleted regions (w_p and w_n) are determined by classical means. Let the metallurgical junction be at position w_j (on Fig. 2 the l.h.s. of the structure is the n-side). At the beginning of each simulation step electrons are injected in the regions

$$0 \leq z < w_j - w_n \quad (2.18)$$

and holes are injected in the (r.h.s.) region

$$L - (w_j + w_p) < z \leq L \quad (2.19)$$

to maintain charge neutrality (i.e. the total charge of carriers in these regions should be equal to the total net charge of dopant ions).

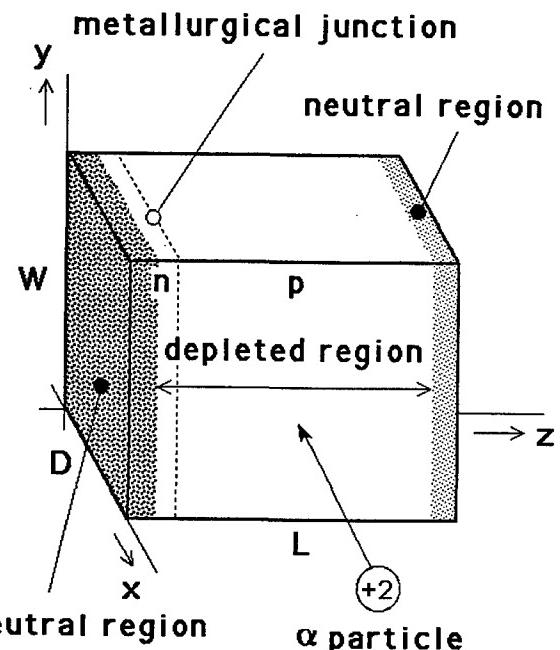


Fig. 2. pn junction structure.

For MOS structures (Fig. 3)

- (a) artificial boundary conditions are applied for the $x - z$ planes at $y = 0$ and at $y = W$.
- (b) Dirichlet boundary conditions are applied for the $y - z$ planes at $x = 0$ assuming zero potential in the non-depleted source region, and a potential calculated by classical means in the depleted source and bulk region, and
- $y - z$ planes at $x = L$ assuming a potential U_{DS} in the non-depleted source region, and a potential calculated by classical means in the depleted source and bulk region, and for
- $x - y$ planes at $z = D$, assuming a potential ϕ_{diff} .

At the beginning of each simulation step, carriers are injected into predefined source, drain and bulk neutral regions in order to maintain charge neutrality.

(c) Partly Dirichlet-, partly Neumann boundary conditions are used on the top of the structure (at $z = 0$).

For the source and drain regions a Dirichlet boundary condition is used with potentials calculated by classical

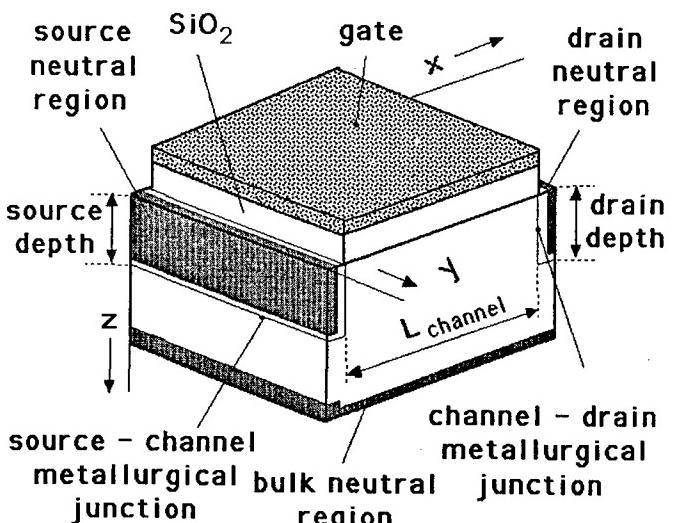


Fig. 3. MOS transistor structure.

means, and Neumann boundary condition is used for the part of the Si-SiO₂ interface, which is under the gate electrode

$$\epsilon_{\text{Si}} \mathbf{grad} \psi(\mathbf{r} + 0) \cdot \bar{\mathbf{n}} = \epsilon_{0x} \mathbf{grad} \psi(\mathbf{r} - 0) \cdot \bar{\mathbf{n}} \quad (2.20)$$

Carriers reaching or passing through the Si-SiO₂ interface are either surface scattered or lost by tunnelling into the oxide layer.

2.4. Scattering processes

The following scattering effects are taken into consideration:

- Coulomb scattering (ionised impurity scattering, scattering on charged interface states and carrier-carrier scatterings);
- Lattice scattering and intervalley scattering;
- Surface scattering.

It is important to note that the majority of the scattering events is deterministic (according to our experiences, more than 90%), therefore the method can be considered as a strongly deterministic simulation method.

2.4.1. Coulomb scattering. Since for each charged particle the exact electric field is analytically determined (only the field of the relatively smooth Laplace potential is evaluated by numerical differentiation), the charged particle interactions are inherently accounted for. This results in the exact evaluation of all Coulomb scatterings (ionised impurity scattering, scattering on charged interface states and carrier-carrier scatterings).

2.4.2. Lattice scattering and intervalley scattering. The interactions between the carriers and the lattice are described by Bragg scattering: If a carrier leaves the boundary of the first Brillouin zone, then it is scattered back inside the first Brillouin zone.

In 2D electron gas layers (e.g. in the inversion layer of MOS structures (f- and g-phonon intervalley scattering is taken into consideration. This model is completed with a corresponding carrier-lattice energy exchange model and with an energy transport model in the crystal lattice [4]. In the bulk region scatterings are modelled quite simply: a thermodynamic approach is applied¹⁰.

2.4.3. Surface scattering. For surface scattering at the Si-SiO₂ interface, which is important for MOS structures, three different phenomena are taken into consideration:

- the Coulomb scattering on charged interface states (see Coulomb scattering),
- elastic surface scattering, if a carrier reaches the Si-SiO₂ interface,
- specular surface scattering, if a carrier reaches the Si-SiO₂ interface.

The ratio between the elastic and specular surface scattering is controlled by the Fuchs parameter [5] and selected by uniformly distributed random numbers. In case of specular scattering a new electron wave vector is generated with

random components. this is the only non-deterministic model in the method.

2.5. Generation-recombination processes

The generation recombination through recombination centres (SRH generation recombination) can be neglected, since the probability of such events during a reasonable simulated time is far below 1.

Under certain circumstances (e.g. when an α particle transition or in the inversion layer of MOS transistors), the impact ionisation and its reverse process, the Auger recombination must be taken into consideration. These play increasingly important role in modern mesoscopic semiconductor devices.

2.5.1. Impact ionisation. Impact ionisation happens if the kinetic energy of a carrier grows over a threshold energy (for silicon the impact ionisation threshold energy is assumed to be $W_{\text{th}} = 1.31 \text{ eV}$ according [6]). When a carrier has an energy of $W > W_{\text{th}}$, a new electron-hole pair is generated. It is assumed, that the ionising particle loses its whole kinetic energy, and an energy equal to the band gap energy W_G is transferred to the lattice.

The momentum and energy conservation yield the conditions

$$W_{\text{total}} = W_{\text{e0}} - W_G = W_e + W_h \quad (2.21)$$

$$\mathbf{p}_{\text{total}} = \mathbf{p}_{\text{e0}} = \mathbf{p}_e + \mathbf{p}_h \quad (2.22)$$

where the momentum of the ionising electron is \mathbf{p}_{e0} , the kinetic energy of the generated electron and hole are W_e and W_h , the momentum of the generated electron and hole are \mathbf{p}_e and \mathbf{p}_h , respectively.

2.5.2. Auger recombination. For the Auger recombination a phenomenological approach is used, based on the scattering cross section concept.

The Auger-recombination rate, according to [7], is given by

$$R = B n^2 p \quad (2.23)$$

where B is the Auger recombination coefficient (theoretical value $3.5 \cdot 10^{-42} \text{ m}^6/\text{s}$, experimental value $2.8 \cdot 10^{-43} \text{ m}^6/\text{s}$). If

$$B n^2 \Delta t \geq 1 \quad (2.24)$$

then one recombination event takes place in an elementary volume of $1/p$. Assuming $n > p$ and a quasi-uniform carrier distribution, the upper limit for the distance between a hole and the nearest electron is

$$\frac{\sqrt{3}}{2} \cdot d_n = \frac{\sqrt{3}}{2} \cdot \frac{1}{n^{1/3}} \quad (2.25)$$

where d_n is the distance between the electrons.

In our approach a recombination event occurs, when the distance sinks below this limit, yielding the critical distance between one electron and one hole

$$d_{eh} = \frac{\sqrt{3}}{2} \cdot (B \Delta t)^{1/6} \quad (2.26)$$

For a time step of $\Delta t = 2.5 \times 10^{-15} \text{ s}$ a critical distance of $d_{eh} = 1.15 \div 1.76 \cdot 10^{-10} \text{ m}$ is obtained.

¹⁰ An electron corresponds to that ellipsoid, where it has minimal energy. Hence, an intervalley scattering occurs for an electron of the i th ellipsoid if the distance in the k -space between the end of the actual wave vector and the centre of the i th ellipsoid is larger than the distance between the end of the actual wave vector and the centre of the j th ellipsoid. If this is the case, the electron leaves the i th ellipsoid and will be associated with the j th ellipsoid in the future. This approach assumes that only the energy is changed, the k -vector is conserved.

3. The ionization process of α particles

The ionization process is caused by an α particle passing through the depleted region of a semiconductor device (e.g. a reverse biased pn-junction).

3.1. The α particle

An α particle is a double ionised He ion, created by the process



where z is the atomic number, m is the atomic weight and Y is an arbitrary element.

The initial velocity of an α particle is [8]

$$v_0 = 1.5 \cdots 2 \times 10^7 \text{ m/s} \quad (3.2)$$

The initial energy of the α particle is given by¹¹

$$\begin{aligned} E_0 &= \frac{m_\alpha v_0^2}{2} \\ &\approx 0.75 \cdots 1.4 \times 10^{-12} \text{ VAs} \\ &\approx 4.6 \cdots 8.8 \text{ MeV} \end{aligned} \quad (3.3)$$

3.2. Interaction with the Si lattice

In the lattice of Si the minimal distance between the Si lattice ions is given by

$$d_{\min} = \frac{a_{\text{Si}}}{\sqrt{2}} \approx 0.378 \text{ nm} \quad (3.4)$$

The path of an α particle, which enters the structure at a point $r_0(x_0, y_0, z_0)$ and leaves at the point $r_1(x_1, y_1, z_1)$, is given by¹²

$$\mathbf{r}(t) = \mathbf{r}_0 + t \cdot (\mathbf{r}_1 - \mathbf{r}_0) \quad (3.5)$$

The t parameter is zero at the entry point and one at the point where the α particle leaves the structure.

The transit time of the particle (taking into consideration that the penetration depth of the α particle in Si is much greater than the path of the particle through the structure) is given by

$$\tau = \frac{|\mathbf{r}_1 - \mathbf{r}_0|}{v_0} \approx 1.25 \cdots 2.9 \times 10^{-14} \text{ s} \quad (3.6)$$

When an α particle approaches Si nuclei for a distance less or equal to $d_{\min}/2$, the α particle in question is scattered and an electron-hole pair is generated. The number of the created electron-hole pairs is given by

$$N_{\text{pairs}} = \frac{|\mathbf{r}_1 - \mathbf{r}_0|}{d_{\min}} \quad (3.7)$$

In a structure having characteristic dimensions of $L = 250$ nm, the electron-hole pair generation induced by a single α particle takes place 1000–1800 times, depending upon the direction of the particle.

Since the scattering of an α particle on a Si nucleus is elastic, only a very small part of the particle energy and momentum is used to create electron-hole pairs. Therefore

an adequate assumption is that the energy and momentum of the created electron and hole should have an average equal to the thermal equilibrium values.

4. Results and discussion

For the effect of α particles in semiconductor devices a reverse biased pn-junction structure has been studied with our simulation method.

4.1. The simulated structure

The simulated structure is the same, as the structure given on Fig. 2. The pn junction formed on $\langle 100 \rangle$ oriented p-type Si substrat has been examined. The homogeneous acceptor doping concentration was $N_A = 2 \times 10^{22} \text{ m}^{-3}$. The length of the structure was $L = 358$ nm, the width $W = 250$ nm and the depth $D = 300$ nm. We assumed a Gaussian donor doping profile, with an initial donor concentration $N_{\text{D},0} = 2 \times 10^{23} \text{ m}^{-3}$ at $x = 0$, and a diffusion depth of $\lambda = 50$ nm. The metallurgical junction was at $x_j = 52.2$ nm.

4.2. Operating point and the α particle

The diode has been reverse biased by -10 V. The response of this structure to a single α particle was simulated. The velocity of the incident α particle was $15 \times 10^6 \text{ m/s}$ (4.7 MeV energy). The particle entered the structure at the time t_0 at position $x_0 = 0.5 L$, $y_0 = 0$, $z_0 = 0.25 D$, and left the structure 0.0197 ps later at point $x_1 = 0.75 L$, $y_1 = W$, $z_1 = 0.75 D$.

4.3. Simulation results

The above described structure was studied by the Micro-MOS 3D particle dynamics Monte Carlo simulation program on a DEC AXP workstation. Here we present some simulation results.

Figure 4 shows the number of electrons and holes vs. time.

Figure 5 shows the carrier distribution in the structure just at the time instant when the α particle entered, and in seven subsequent time instants. Along the path of the α particle, in its very narrow neighbourhood a dense electron-hole plasma is formed. The maximal local carrier concentrations has been found as high as $5.56 \times 10^{25} \text{ m}^{-3}$.

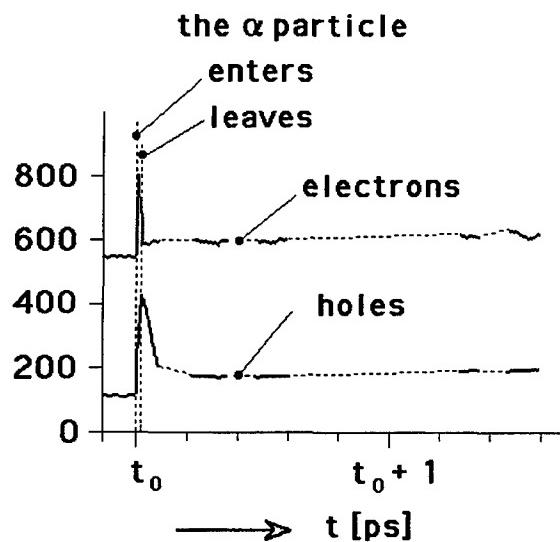


Fig. 4. The number of electrons and holes vs. time.

¹¹ The mass of the α particle is 4.00390 times of the proton mass. Using the value of $m_{\text{proton}} = 1836.1388 m_0$ yields $m_\alpha = 6.697 \times 10^{-27} \text{ kg}$.

¹² For simplicity, the path of the particle is assumed to be a straight line.

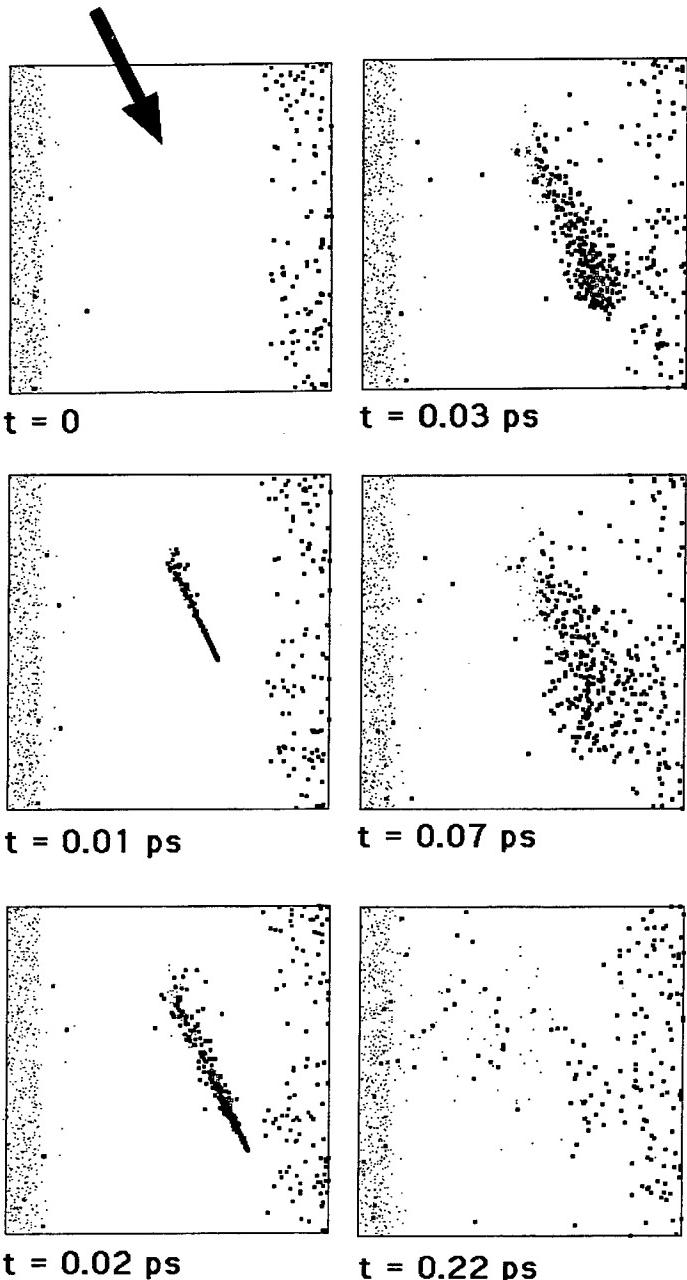


Fig. 5. Carrier distribution in the structure before, during and after the α -particle ionisation process. The dots represent the electrons, while \square the holes.

It can be observed, that during the transition, and later for a time approx. 0.1 ps the very strong Auger recombination uses up a sufficient part of the generated carriers (in our example from the generated 648 pairs about 590 pairs are consumed by Auger recombination during a time of 10^{-13} s). However, the remaining part of the generated electron-hole pairs is sufficient to initialise an avalanche effect in the depleted region by impact ionisation. This can

be also observed on the Fig. 4: the spike near the beginning of the simulation corresponds to the formation of the electron-hole plasma and later the increased number of carriers corresponds to the increased reverse current due to the avalanche mechanism.

4.4. Conclusions

(1) Comparing our results to the results published at the same time for a nearby similar problem presented by Moglestue *et al.* [9], it can be observed that the maximal plasma density given by the two methods are in an acceptable agreement, but we do not agree with the statement in [9] that "The kinetic energy ... is insufficient to cause further impact ionisation ...". The method described in [9] is two dimensional and is based on charge assignment and on the solution of the Poisson equation, therefore an averaged field acts on the carriers. Our method considers the very chaotic local field acting to carriers, therefore the carriers' energy gained from the field may reach to such a high level, which is enough for a quite strong impact ionisation.

(2) It has been demonstrated, that our method is a successful tool for studying and better understanding various phenomena in semiconductor structures, giving insight into the details of different processes and their time dependence¹³.

References

1. Kocsis, T., Tarnay, K., Masszi, F. and Poppe, A., "Modeling the 2D Behaviour of Electrons near the Si-SiO₂ Interface by a 3D Particle Dynamics Monte Carlo Simulator". Accepted by Physics Computing, PC'94 (Lugano, Switzerland, 21–28 Aug 1994).
2. Ivey, J.L. and Mieher, R. L., Phys. Rev. Lett., **29**, pp. 176–178 (1972).
3. Hockney, R. W. and Eastwood, J. W., "Computer Simulation Using Particles" (Adam Hilger, Bristol and Philadelphia 1988).
4. Tarnay, K., Habermajer, I., Poope, A., Kocsis, T. and Masszi, F., "Modeling the Carrier – Lattice Interactions and the Energy Transport in a 3D Particle Dynamics Monte Carlo Simulator for MOS Structures". Abstracts of NASECODE X, 21–24 June 1994, Dublin, 28–29¹⁴.
5. Fuchs, K., Proc. Cambr. Phil. Soc. **34**, 1938.
6. Quade, W., Shöll, E. and Rudan, M., Solid State Electronics, **36** (10), 1493–1505, (1993).
7. Dziewior, J. and Schmid, W., Appl. Phys. Lett. **31** 346–348 (1977).
8. Kitaigorodsky, A., "Introduction to Physics" (Foreign Languages Publishing House, Moscow 1978), p. 541.
9. Moglestue, C., Buot, F., and Anderson, W. T., "Thermally induced Failure in GaAs Transistors Exposed to Alpha Particle Irradiation." Abstracts of NASECODE X, 21–24 June 1994, Dublin, 17–10¹⁴.

¹³ At the 16th Nordic Semiconductor Meeting (Laugarvatn, Iceland, June 12–15, 1994) a "computer movie" has been presented, showing the real space and wave vector space behaviour of electrons and holes during and after the transition of an α particle.

¹⁴ Will be published in details in the International Journal for Computation and Mathematics in Electrical and Electronic Engineering (COMPEL).

A Theoretically Accurate Mobility Model for Semiconductor Device Drift-Diffusion Simulation

E. Velmre,¹ A. Udal,¹ T. Kocsis² and F. Masszi²

¹ Institute of Electronics, Tallinn Technical University, Akadeemia tee 1, EE0026 Tallinn, Estonia

² Scanner Lab, Inst. of Technology, Dept. of Electronics, Uppsala University, P.O. Box 534, S-75121 Uppsala, Sweden

Received May 4, 1994; accepted June 14, 1994

Abstract

A novel semiconductor charge carrier mobility model obtained by Kohler's variational method for Boltzmann transport equation solution is presented. Acoustic phonon scattering, ionized impurity scattering, and carrier-carrier scattering are taken into account. Majority electron and hole mobilities in n- and p-type silicon versus impurity and carrier concentration, and temperature are calculated and compared with published experimental data.

1. Introduction

If the momentum exchange between electrons and holes due to electron-hole scattering (EHS), i.e. the carrier drag effect [1] is taken into account Van Roosbroeck's drift-diffusion transport equations [2] have to be extended with cross terms. Thus, for isothermal case [3]

$$J_e = -qn_e \mu_{e1} \nabla \psi + k_B T \mu_{e2} \nabla n_e + k_B T \mu_{e3} \nabla n_h \quad (1)$$

$$J_h = -qn_h \mu_{h1} \nabla \psi - k_B T \mu_{h2} \nabla n_h - k_B T \mu_{h3} \nabla n_e \quad (2)$$

Here the electron and hole mobility-type transport coefficients influenced by EHS differ from mobilities in the commonly used Van Roosbroeck transport equations. Transport coefficients μ_{e1} and μ_{h1} may be considered as drift mobilities of electron and holes, respectively. Coefficients μ_{e2} and μ_{h2} may be named as "diffusion mobilities" when μ_{e3} and μ_{h3} are "drag mobilities". In case of negligible EHS effect $\mu_{e3} = \mu_{h3} = 0$, $\mu_{e1} = \mu_{e2} = \mu_{e0}$, and $\mu_{h1} = \mu_{h2} = \mu_{h0}$, where μ_{e0} and μ_{h0} are mobilities not influenced by EHS. As result, eqs (1), (2) coincide with Van Roosbroeck's equations, where mobilities and diffusion coefficients are interrelated by Einstein's relationship.

Drift mobilities μ_{e1} and μ_{h1} may be represented as linear combinations of diffusion and drag mobilities:

$$\begin{aligned} \mu_{e1} &= \mu_{e2} - \mu_{e3} = \mu_{e2} - \frac{n_h}{n_e} \mu_{e3}, \\ \mu_{h1} &= \mu_{h2} - \mu_{e3} = \mu_{h2} - \frac{n_e}{n_h} \mu_{e3} \end{aligned} \quad (3)$$

This choice of independent mobilities may be supported also by thermodynamical argumentation. For that, substituting (3) into (1), (2) we obtain carrier transport equations in the form derived from principles of irreversible thermodynamics [4, 5]:

$$J_e = \sigma_e (-\nabla \phi_e) - \sigma_{eh} (-\nabla \phi_h) \quad (4)$$

$$J_h = -\sigma_{he} (-\nabla \phi_e) + \sigma_h (-\nabla \phi_h) \quad (5)$$

where the quasi-Fermi potential gradients are

$$-\nabla \phi_e = -\nabla \psi + \frac{k_B T}{q} \cdot \frac{1}{n_e} \nabla n_e,$$

$$-\nabla \phi_h = -\nabla \psi - \frac{k_B T}{q} \cdot \frac{1}{n_h} \nabla n_h$$

and the conductivity-type transport coefficients are

$$\sigma_e = q n_e \mu_{e2}, \quad \sigma_h = q n_h \mu_{h2}, \quad \sigma_{eh} = q n_h \mu_{e3},$$

and

$$\sigma_{he} = q n_e \mu_{h3}$$

Non-zero quantities σ_{eh} and σ_{he} have been arised from EHS effect and are introducing direct coupling between the electron and hole currents. In accordance with Onsager's relationship $\sigma_{eh} = \sigma_{he}$.

In general, the transport coefficients σ_e , σ_h , σ_{eh} and σ_{he} are tensors of rank 2. However, for crystals of cubic symmetry these tensors can be transformed into diagonal form with identical components, and therefore can be considered as scalar quantities. This is valid for mobilities $\mu_{e1} \dots 3$ and $\mu_{h1} \dots 3$ as well.

To use transport eqs (1), (2) or (4), (5) for semiconductor simulation purposes the relevant transport coefficients have to be specified. In Section 2, we represent a method of mobilities calculation based on the Kohler's variational principle [6, 7]. In Section 3, a comparison of calculated and measured mobility values is presented.

2. Mobility model

Kohler's variational method for Boltzmann transport equation (BTE) solution enables the inclusion of all essential scattering mechanisms without any relaxation time assumption. In this paper, the following scattering mechanisms will be considered:

- longitudinal acoustic phonon scattering, i.e. the ordinary lattice scattering for electrons (ELS) and holes (HLS);
- ionized impurity scattering for electrons (EIS) and holes (HIS);
- electron-electron (EES) and hole-hole scattering (HHS);
- electron-hole (EHS) and hole-electron scattering (HES).

Mobilities in eqs (1) and (2) obtained using Kohler's method can be written as [8] for electrons

$$\mu_{e2}^{(N)} = \frac{q}{m_e} \cdot \frac{1}{|G|} \sum_{r=0}^N \frac{(r+3/2)!}{(3/2)!} \Delta_{2r+2, 1}^{(E)}$$

$$\mu_{e3}^{(N)} = -\frac{q}{m_e} \cdot \frac{1}{|G|} \sum_{r=0}^N \frac{(r+3/2)!}{(3/2)!} \Delta_{2r+3,1}^{(E)}$$

$$\mu_{e1}^{(N)} = \mu_{e2}^{(N)} - \frac{n_h}{n_e} \mu_{e3}^{(N)}$$

and for holes

$$\mu_{h2}^{(N)} = -\frac{q}{m_h} \cdot \frac{1}{|G|} \sum_{r=0}^N \frac{(r+3/2)!}{(3/2)!} \Delta_{2r+3,1}^{(E)}$$

$$\mu_{h3}^{(N)} = \frac{q}{m_h} \cdot \frac{1}{|G|} \sum_{r=0}^N \frac{(r+3/2)!}{(3/2)!} \Delta_{2r+2,1}^{(E)}$$

$$\mu_{h1}^{(N)} = \mu_{h2}^{(N)} - \frac{n_e}{n_h} \mu_{h3}^{(N)}$$

Here, $z! = \Gamma(z+1) = z\Gamma(z)$ and $N = 0, 1, 2, \dots, \infty$ is the BTE variational solution approximation order. Quantities m_e^* , m_h^* are the relative effective masses of electrons and holes, respectively.

Determinants $\Delta_{2r+2,1}^{(E)}$, $\Delta_{2r+3,1}^{(E)}$ and $|G|$ can be derived from determinants $|D_\alpha^{(E)}|$ introduced by Meyer [9]. However, the major importance in mobility calculations has determinant $|G|$, which elements contain all necessary information on scattering parameters:

$$|G| = \begin{vmatrix} * \Gamma_{00}^{ee} & * \Gamma_{00}^{eh} & \cdots & * \Gamma_{0N}^{ee} & * \Gamma_{0N}^{eh} \\ * \Gamma_{00}^{he} & * \Gamma_{00}^{hh} & \cdots & * \Gamma_{0N}^{he} & * \Gamma_{0N}^{hh} \\ \vdots & \vdots & G & \vdots & \vdots \\ * \Gamma_{NO}^{ee} & * \Gamma_{NO}^{eh} & \cdots & * \Gamma_{NN}^{ee} & * \Gamma_{NN}^{eh} \\ * \Gamma_{NO}^{he} & * \Gamma_{NO}^{hh} & \cdots & * \Gamma_{NN}^{he} & * \Gamma_{NN}^{hh} \end{vmatrix}$$

Elements in the principal diagonals of 2×2 sub-blocks in $|G|$ are sums of terms created by certain scattering mechanism including carrier-carrier scattering effects:

$$*\Gamma_{rs}^{ee} = *d_{rs}^{e-ac} + *d_{rs}^{e-i} + *d_{rs}^{e-e} + *g_{rs}^{e-h},$$

$$*\Gamma_{rs}^{hh} = *d_{rs}^{h-ac} + *d_{rs}^{h-i} + *d_{rs}^{h-h} + *g_{rs}^{h-e}.$$

EHS and HES give two types of scattering terms in $|G|$ elements. Terms $*g_{rs}^{\alpha-\beta}$ enter into the principal diagonal elements of sub-blocks. The off-diagonal elements of sub-blocks contain only $*h_{rs}^{\alpha-\beta}$ terms, which are responsible for carrier drag effect description:

$$*\Gamma_{rs}^{eh} = *h_{rs}^{e-h}, \quad *\Gamma_{rs}^{he} = *h_{rs}^{h-e}.$$

If EHS and HES are negligible, then $*\Gamma_{rs}^{eh} = *\Gamma_{rs}^{he} = 0$.

Determinants $|D_\alpha^{(E)}|$ for electrons ($\alpha = e$) and holes ($\alpha = h$) may be regarded as bordered determinants formed by bordering the determinant $|G|$ with a prefixed row and column.

There is the first row for $\alpha = e$

$$(0 \ * \beta_{e0}^{(E)} \ 0 \ * \beta_{e1}^{(E)} \ 0 \ \cdots \ * \beta_{eN}^{(E)} \ 0)$$

and for $\alpha = h$

$$(0 \ 0 \ * \beta_{h0}^{(E)} \ 0 \ * \beta_{h1}^{(E)} \ \cdots \ 0 \ * \beta_{hN}^{(E)})$$

and the first column for $\alpha = e$

$$(0 \ * \beta_{e0}^{(E)} \ -R_{e,h} * \beta_{h0}^{(E)} * \beta_{e1}^{(E)} \ -R_{e,h} * \beta_{h1}^{(E)} \ \cdots \ * \beta_{eN}^{(E)} \ -R_{e,h} * \beta_{hN}^{(E)})'$$

and for $\alpha = h$

$$(0 \ R_{h,e} * \beta_{e0}^{(E)} \ -* \beta_{h0}^{(E)} \ R_{h,e} * \beta_{e1}^{(E)} \ -* \beta_{h1}^{(E)} \ \cdots \ R_{h,e} * \beta_{eN}^{(E)} \ -* \beta_{hN}^{(E)})'$$

where the sign ' denotes transposition. $R_{e,h}$ and $R_{h,e}$ are certain scalar quantities depending on electron and hole quasi-Fermi potential gradients.

The determinant $\Delta_{2r+2,1}^{(E)}$ is a minor of the determinant $|D_\alpha^{(E)}|$ obtained by omitting the $(2r+2)$ th row and the 1st column, without giving the sign $(-1)^{2r+3}$.

Whereas the first column of $|D_\alpha^{(E)}|$ is always omitted by minor calculation quantities $R_{e,h}$ and $R_{h,e}$ do not occur in mobility calculations.

To facilitate calculations all determinant elements are scaled down by the factor

$$\frac{n_e + n_h}{n_e} d_{00}^{e-ac}(300 \text{ K}),$$

where $d_{00}^{e-ac}(300 \text{ K})$ is the zeroth-order scattering element for electron-acoustic phonon interactions at $T = 300 \text{ K}$

$$d_{00}^{e-ac}(300 \text{ K}) = \frac{32}{9\pi} \cdot \frac{qn_e}{m_e^2(300 \text{ K})\mu_{el}(300 \text{ K})}$$

where $\mu_{el}(300 \text{ K})$ is the electron-acoustic phonon scattering mobility at $T = 300 \text{ K}$.

Scaled scattering terms are marked by an asterisk, e.g.

$$*d_{rs}^{\alpha-ac} \equiv d_{rs}^{\alpha-ac} / \left(\frac{n_e + n_h}{n_e} d_{00}^{e-ac}(300 \text{ K}) \right).$$

As a consequence, all determinant elements $*\Gamma_{rs}^{\alpha\beta}$ are dimensionless as well. Scaled terms for different scattering mechanisms are given in Appendix.

By Kohler's method mobilities as transport parameters can be calculated in the zeroth, first, second, and higher approximations:

$$\mu_{ai}^{(0)}, \mu_{ai}^{(1)}, \mu_{ai}^{(2)}, \dots, \mu_{ai}^{(N)}$$

where N is the order of approximation, α is e for electrons and h for holes, and $i = 1, 2, 3$.

The case $N \rightarrow \infty$ corresponds to the exact solution of the BTE. Successive approximations for mobility and electrical conductivity show rapid convergence after the first order correction [10]. Often the first and even the zeroth approximation provides practically acceptable accuracy of the results.

3. Calculation of mobilities

It is of considerable interest to apply described mobility model to calculate, first of all, electron and hole mobilities in silicon versus impurity and carrier concentration, and temperature. Obtained results are compared with the experimental data from papers by Klaassen [11], Dannhäuser [12], and Krausse [13].

To obtain results, presented below, for basic mobilities $\mu_{el}(300 \text{ K})$, $\mu_{hl}(300 \text{ K})$ the Scharfetter-Gummel [15] data 1400 and $480 \text{ cm}^2/\text{V}\cdot\text{s}$ were used. The temperature-dependence exponents for these two mobilities -2.285 and -2.247 were taken by Klaassen [11]. All adjustment coefficients introduced in Appendix were not used, i.e. the unit values were assigned to these parameters. Electron and hole effective masses temperature dependences were calculated accordingly to Vankemmel *et al.* [14], obtaining, for

instance, $m_e^* = 1.185$ and $m_h^* = 1.153$ at $T = 300$ K. Calculations were performed for zeroth and first approximation order.

Fig. 1 represents the majority electron mobility temperature dependence for 5 different donor doping levels. Fig. 2 gives the similar plot for majority hole mobility for 5 different acceptor doping levels. In these calculations the low injection level quasineutral situation was assumed and thereby nearly equal to doping majority carrier concentration was present. As may be seen, a reasonable agreement between the non-adjusted model and experimental data exists, but the need of some model parameter adjustment, especially for electrons, is obvious as well. In the present work this adjustment was intentionally avoided to reveal the novel model behaviour in a more clear way. It should be noted that the present model gives the experimentally verified maximum point on mobility versus temperature curves at low temperatures and high doping levels.

Fig. 3 compares the calculated sum of electron and hole mobilities with the well-known Dannhäuser-Krausse

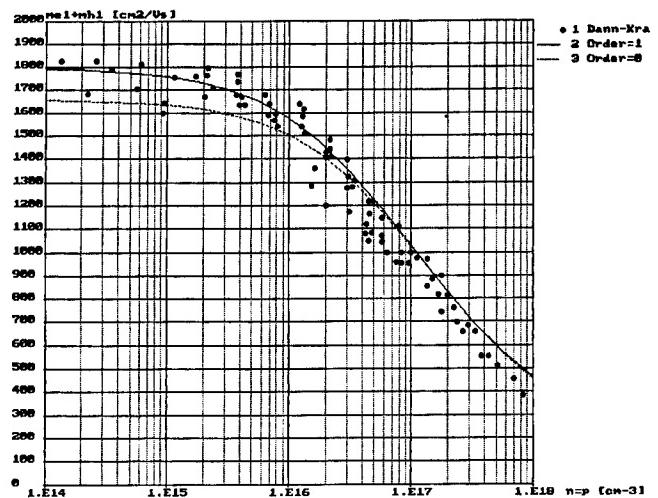


Fig. 3. Sum of electron and hole mobilities versus carrier concentration for low-doped silicon.

Solid line: model, order 1

Dotted line: model, order 0

Points: experimental data of Dannhäuser-Krausse [12, 13]

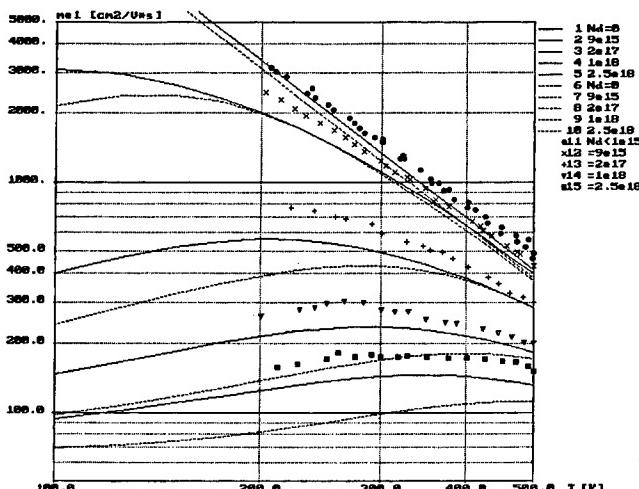


Fig. 1. Majority electron mobility versus temperature for different donor doping concentrations.

Solid lines: model, order 1

Dotted lines: model, order 0

Points: experimental data from Klaassen's paper [11]

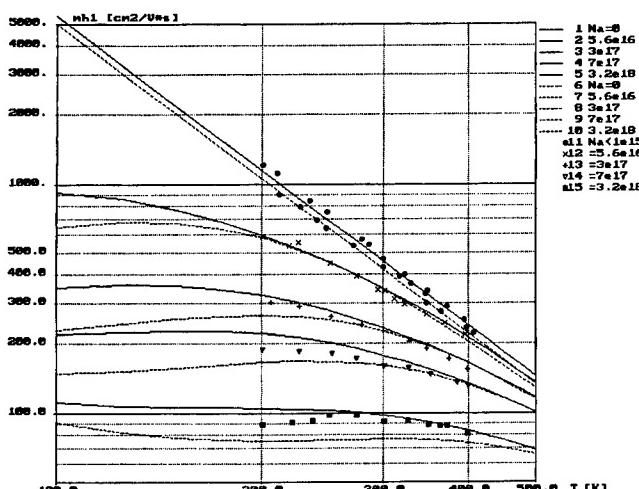


Fig. 2. Majority hole mobility versus temperature for different acceptor doping concentrations.

Solid lines: model, order 1

Dotted lines: model, order 0

Points: experimental data from Klaassen's paper [11]

experimental data [12, 13], describing summary mobility reduction in low-doped silicon at high carrier densities. As one can see, the present model without any adjustment yields a good agreement with these data.

4. Conclusions

Comparison of calculated and measured mobility values in Section 3 proves that our novel mobility model has a fairly good exactness in wide range of doping concentration and temperature in silicon. It is worth to stress that mentioned calculations have been performed without any parameter adjustment. For further improvement of the mobility model, an adjustment of correction factors relevant to the shielding constants seem to be appropriate. The zeroth and first approximation order mobility model is implemented in the semiconductor device simulator DYNAMIT, developed at Tallinn Technical University. An extension of the model to higher approximation order is under development.

Acknowledgement

This research has been supported in part by the Estonian Science Foundation and the Swedish Institute.

Appendix

In the following, effective isotropic parabolic bands for both electrons and holes are assumed. Electron and hole scattering cross-sections have been used in the first Born approximation.

1. Longitudinal acoustic phonon scattering terms

Terms for electrons ($\alpha = e$) and holes ($\alpha = h$) are

$$*d_{rs}^{\alpha-ac} = a_{\alpha-ac} \frac{(r+s+2)!}{2!} \frac{n_\alpha}{n_e + n_h} \frac{\mu_{eL}(300 \text{ K})}{\mu_{\alpha L}} \left(\frac{m_e^*(300 \text{ K})}{m_\alpha^*} \right)^2$$

where $\mu_{\alpha L}$ may be approximated by the lattice scattering mobilities, measured in a high-purity semiconductor. Scharfetter and Gummel [15] have suggested the values $\mu_{eL}(300 \text{ K}) = 1400$ and $\mu_{hL}(300 \text{ K}) = 480 \text{ cm}^2/(\text{V} \cdot \text{s})$. The quantities m_α^* are the relative (i.e. dimensionless) effective masses $m_\alpha^* = m_\alpha/m_0$ of electrons or holes. Dimensionless correction factors $a_{\alpha-ac}$ serve for fitting of calculated and measured mobility values.

The temperature dependence of mobilities $\mu_{\alpha L}$ is conventionally approximated by

$$\mu_{\alpha L} = \mu_{\alpha L}(300 \text{ K}) \left(\frac{T}{300 \text{ K}} \right)^{-\alpha_{T\mu\alpha L}}$$

where $\alpha_{T\mu\alpha L} = 2.285$ and $\alpha_{T\mu hL} = 2.247$ [11].

2. Ionized impurity scattering terms

Terms for electrons ($\alpha = e$) and holes ($\alpha = h$) are

$$\begin{aligned} *d_{rs}^{\alpha-i} &= a_{\alpha-i} 2.565 \times 10^{-21} (m_e^*(300 \text{ K}))^{1/2} \left(\frac{\varepsilon}{11.7} \right)^{-2} \\ &\quad \times \left(\frac{T}{300 \text{ K}} \right)^{-3/2} \left(\frac{m_e^*(300 \text{ K})}{m_\alpha^*} \right)^{3/2} \\ &\quad \times \mu_{eL}(300 \text{ K}) \frac{n_\alpha N_i}{n_e + n_h} M_{r+s}(\delta_{\alpha-i}^2) \end{aligned}$$

where N_i is the total impurity concentration ($N_i = N_D^+ + N_A^-$) in cm^{-3} and the mobility μ_{eL} is in $\text{cm}^2/(\text{V} \cdot \text{s})$.

The parameter integrals M_{r+s} are defined in [10] as

$$\begin{aligned} M_{r+s}(\delta_{\alpha-i}^2) &= \int_0^\infty \exp(-x^2) x^{2(r+s)+1} \\ &\quad \times \left[\ln \left(1 + \frac{4x^2}{\delta_{\alpha-i}^2} \right) - \frac{1}{1 + \delta_{\alpha-i}^2/(4x^2)} \right] dx, \end{aligned}$$

where the dimensionless parameter $\delta_{\alpha-i}^2$ which is related to the shielding constant of an ionic Coulomb potential, is determined by

$$\begin{aligned} \delta_{\alpha-i}^2 &= b_{\alpha-i} 4.405 \times 10^{-20} m_\alpha^{*-1} \left(\frac{\varepsilon}{11.7} \right)^{-1} \\ &\quad \times \left(\frac{T}{300 \text{ K}} \right)^{-2} (n_e + n_h) \end{aligned}$$

where n_α is in cm^{-3} . Dimensionless correction factors $a_{\alpha-i}$ and $b_{\alpha-i}$ serve for fitting of calculated and measured mobility values.

3. Electron-electron and hole-hole scattering terms

The formulas for EES and HHS terms are similar to the carrier-ion scattering terms description. Thus, for electrons ($\alpha = e$) and holes ($\alpha = h$)

$$\begin{aligned} *d_{rs}^{\alpha-\alpha} &= a_{\alpha-\alpha} 9.070 \times 10^{-22} (m_e^*(300 \text{ K}))^{1/2} \\ &\quad \times \left(\frac{\varepsilon}{11.7} \right)^{-2} \left(\frac{T}{300 \text{ K}} \right)^{-3/2} \left(\frac{m_e^*(300 \text{ K})}{m_\alpha^*} \right)^{3/2} \\ &\quad \times \mu_{eL}(300 \text{ K}) \frac{n_\alpha^2}{n_e + n_h} P_{rs}(\delta_{\alpha-\alpha}^2) \end{aligned}$$

where

$$P_{0s} = P_{r0} = 0, \quad \text{if } r, s = 0, 1, 2, \dots$$

$$P_{11}(\delta_{\alpha-\alpha}^2) = L_1(\delta_{\alpha-\alpha}^2).$$

The parameter integral is defined by the formula

$$\begin{aligned} L_n(\delta_{\alpha-\alpha}^2) &= \int_0^\infty \exp(-x^2) x^{2n+1} \\ &\quad \times \left\{ \left[5 \left(1 + \frac{\delta_{\alpha-\alpha}^2}{x^2} \right) - \frac{1}{1 + \delta_{\alpha-\alpha}^2/x^2} \right] \right. \\ &\quad \left. \times \ln \left(1 + 2 \frac{x^2}{\delta_{\alpha-\alpha}^2} \right) - 10 \right\} dx, \end{aligned}$$

where the dimensionless parameter

$$\delta_{\alpha-\alpha}^2 = b_{\alpha-\alpha} 8.810 \times 10^{-20} \left(\frac{\varepsilon}{11.7} \right)^{-1} \left(\frac{T}{300 \text{ K}} \right)^{-2} \frac{n_e + n_h}{m_\alpha^*}$$

and n_α is in cm^{-3} .

4. Electron-hole and hole-electron scattering terms

Here, the principal role is playing the term

$$*g_{00}^{e-h} = J^{eh} \frac{n_e n_h}{n_e + n_h} \mu_{eL}^{(0)}(300 \text{ K}),$$

where

$$\mu_{eL}^{(0)} = \frac{9\pi}{32} \mu_{eL}$$

The EHS parameter J^{eh} introduced by Meyer [9] may be calculated as

$$\begin{aligned} J^{eh} &= a_{e-h} 2.902 \times 10^{-21} m_{red}^{*1/2} \left(\frac{\varepsilon}{11.7} \right)^{-2} \\ &\quad \times \left(\frac{T}{300 \text{ K}} \right)^{-3/2} \left(\frac{m_e^*(300 \text{ K})}{m_e^*} \right)^2 J_0, \text{ V} \cdot \text{cm} \cdot \text{s} \end{aligned}$$

where the carrier reduced mass is expressed as

$$m_{red}^* = \frac{m_e^* m_h^*}{M^*}, \quad M^* = m_e^* + m_h^*.$$

The other g -terms are defined as in [16]:

$$*g_{00}^{h-e} = \left(\frac{m_e^*}{m_h^*} \right)^2 *g_{00}^{e-h}$$

$$*g_{01}^{\alpha-\beta} = *g_{10}^{\alpha-\beta} = *g_{00}^{\alpha-\beta} \left(\frac{m_\alpha^*}{M^*} \right) \left[\frac{5}{2} + \left(\frac{m_\beta^*}{m_\alpha^*} \right) \frac{I_1}{J_0} \right]$$

$$\begin{aligned} *g_{11}^{\alpha-\beta} &= *g_{00}^{\alpha-\beta} \left(\frac{m_\alpha^*}{M^*} \right)^2 \\ &\quad \times \left[\frac{55}{4} + 5 \left(\frac{m_\beta^*}{m_\alpha^*} \right) \frac{J_1}{J_0} + \left(\frac{m_\beta^*}{m_\alpha^*} \right)^2 \frac{J_2}{J_0} + 2 \left(\frac{m_\beta^*}{m_\alpha^*} \right) \frac{I_1}{J_0} \right] \end{aligned}$$

The h -terms are also defined following [16]:

$$*h_{00}^{e-h} = *h_{00}^{h-e} = - \left(\frac{m_e^*}{m_h^*} \right) *g_{00}^{e-h},$$

$$*h_{01}^{\alpha-\beta} = *h_{10}^{\alpha-\beta} = - \left(\frac{m_\beta^*}{m_\alpha^*} \right) *g_{10}^{\alpha-\beta}$$

$$\begin{aligned} *h_{11}^{e-h} &= *h_{11}^{h-e} = -\left(\frac{m_e^*}{M^*}\right)^2 \\ &\times *g_{00}^{e-h} \left[\frac{55}{4} + \frac{5}{2} \left(\frac{m_e^*}{m_h^*} + \frac{m_h^*}{m_e^*} \right) \frac{J_1}{J_0} + \frac{J_2}{J_0} - 2 \frac{I_1}{J_0} \right] \end{aligned}$$

The parameter integrals are defined by

$$\begin{aligned} J_n(\delta^2) &= \int_0^\infty \exp(-x^2) x^{2n+1} \\ &\times \left[\ln \left(1 + \frac{2x^2}{\delta^2} \right) - \frac{1}{1 + \delta^2/(2x^2)} \right] dx, \end{aligned}$$

$$\begin{aligned} I_n(\delta^2) &= \int_0^\infty \exp(-x^2) x^{2n+1} \\ &\times \left[2 \left(1 + \frac{\delta^2}{x^2} \right) \ln \left(1 + \frac{2x^2}{\delta^2} \right) - 4 \right] dx, \end{aligned}$$

where the dimensionless parameter

$$\begin{aligned} \delta^2 &\equiv \delta_{e-h}^2 = \delta_{h-e}^2 = b_{e-h} 4.405 \times 10^{-20} \\ &\times \left(\frac{\varepsilon}{11.7} \right)^{-1} \left(\frac{T}{300 \text{ K}} \right)^{-2} \frac{n_e + n_h}{m_{red}^*} \end{aligned}$$

and n_α is in cm^{-3} .

5. Carrier concentration related terms

These terms enter into the elements of prefixed rows and columns of the $|G|$ -determinant bordering.

In case of Fermi-Dirac statistics the non-scaled terms are [9]:

$$\beta_{\alpha r}^{(E)} = \frac{(r + 3/2)!}{(3/2)!} \cdot \frac{\mathcal{F}_{r+1/2}(\eta_\alpha)}{\mathcal{F}_{1/2}(\eta_\alpha)} \cdot \frac{n_\alpha}{m_\alpha}$$

The Fermi integrals are defined by

$$\mathcal{F}_j(\eta_\alpha) = \frac{1}{j!} \int_0^\infty \frac{y^j dy}{\exp(y - \eta_\alpha) + 1},$$

where η_α is the reduced Fermi energy for an α -type charge carrier.

For the Maxwell-Boltzmann statistics the scaled terms may be written as

$$*\beta_{\alpha r}^{(E)} = \frac{(r + 3/2)!}{(3/2)!} \cdot \frac{m_e^*(300 \text{ K})}{m_\alpha^*} \cdot \frac{n_\alpha}{n_e + n_h} \frac{m_e(300 \text{ K}) \mu_{eL}^{(0)}(300 \text{ K})}{q}.$$

They have dimensions of time in seconds.

References

1. Paige, E. G. S., J. Phys. Chem. Solids **16**, 207 (1960).
2. Van Roosbroeck, W. V., Bell Syst. Techn. J. **29**, 560 (1950).
3. Velimre, E., Koel, A. and Masszi, F., "Simulation of Semiconductor Devices and Processes" (Edited by S. Selberherr, H. Stippel and E. Strasser) (Springer-Verlag, Wien, New York 1993), vol. 5, p. 433.
4. Kane, D. E. and Swanson, R. M., J. Appl. Phys. **72**, 529 (1992).
5. Lindefelt, U., J. Appl. Phys. **75**, 958 (1994).
6. Kohler, M., Z. Phys. **124**, 772 (1948).
7. Kohler, M., Z. Phys. **125**, 679 (1949).
8. Velimre, E., Trans. of Tallinn Techn. Univ. No. 674, 166 (1988).
9. Meyer, J. R., Phys. Rev. **B21**, 1554 (1980).
10. Appel, J., Phys. Rev. **122**, 1760 (1961).
11. Klaassen, D. B. M., Solid-State Electron. **35**, 961 (1992).
12. Dannhäuser, F., ibid. **15**, 1371 (1972).
13. Krausse, J., ibid. **15**, 1377 (1972).
14. Vankemmel, R., Schoenmaker, W. and De Meyer, K., ibid. **36**, 1379 (1993).
15. Scharfetter, D. L. and Gummel, H. K., IEEE Trans. Electron Devices, **ED-16**, 64 (1969).
16. Appel, J., Phys. Rev. **125**, 1815 (1962).

Different Methods of Noise Reduction in Monte Carlo Simulations of a Schottky Diode

Hans-Erik Nilsson^{1,2}, Ulf Sannemo², C. Sture Petersson¹

¹ Department of Solid State Electronics, The Royal Institute of Technology, Electrum 229, S-164 40 Kista, Sweden

² School of Science and Engineering, Mid-Sweden University S-851 70 Sundsvall, Sweden

Received June 6, 1994; accepted June 14, 1994

Abstract

Methods of decreasing the noise and increasing the computational efficiency in Monte Carlo simulations of semiconductor devices are investigated. A lookup table approach to the charge assignment to mesh points has been implemented which is independent of the complexity of the assignment function. Using this approach reduces the CPU time of the charge assignment to about one third. Potential fluctuations in low field regions are compared for different assignment functions, which shows the advantage of using more complex schemes than those generally used. Results from a full band Monte Carlo simulation of a submicron Schottky diode are presented and the values of the current density for different assignment schemes are compared, showing a lower noise for the higher order schemes. Statistical enhancements by splitting of superparticles in the depletion region has been investigated. The combination of a high order assignment scheme and statistical enhancement by splitting decreases simulation runtimes considerable for a given noise tolerance.

1. Introduction

Monte Carlo (MC) methods have become a standard tool for studying semiconductor materials and devices [1, 2]. Recently several models using realistic band structures have been described and applied to bulk material and devices [3, 4]. An obvious limitation is the computational effort needed to perform this kind of simulation. It is therefore of interest to find ways of making the simulations more efficient. In devices with low-field regions using MC methods in the whole device is inefficient. Methods combining a drift-diffusion (DD) simulator with MC have been presented [5, 6]. Even though these methods reduce the CPU time, the MC part is still time consuming. MC simulations of normally off MESFET structures and Permeable Base Transistors (PBT) demand accurate treatment of a Schottky contact under forward biasing using a small number of superparticles. In MC simulations of a submicron Schottky diode several problems are encountered. In the undepleted region the field is very low and sensitive to fluctuations in the charge density. These fluctuations depend on the number of particles and on how the actual charge to mesh assignment is done. Usually simple schemes like nearest grid point (NGP) and cloud-in-cell (CIC) have been used [7]. In [7, 8] the concept of *aliasing* is discussed extensively and it is shown that its effect on the charge density can be reduced by a proper choice of assignment function. Aliasing is well known from signal theory and several “window functions”, analogous to the charge assignment functions in our case, have been proposed.

In this paper we have studied a submicron Schottky diode with a full band MC method. Of course the full band approach is not necessary or even preferable if one wants to study transport in this particular device only. However, a

device such as the PBT contains forward-biased Schottky contacts near a region where the electric field is high and consequently the use of a full band method would be justified. The purpose of this study is twofold: To examine how reliable the results are when a relatively small number of particles is used and to study the effect of using different charge assignment schemes.

The paper is organized as follows.

In Section 2 the Monte Carlo model and the system that we study are described, Section 3 gives results of the study of potential fluctuations for different assignment functions as well as of using different functions in the same simulation, and a description of the lookup table approach, Section 4 contains MC results for a Schottky diode with different charge assignment functions and in Section 5 we discuss a possible way of handling scarcely populated “rare” regions. Conclusions are given in Section 6.

2. The Monte Carlo model

The MC program we use is a modified version of the full band MC program for bulk Si published in [9]. Several modifications have been made:

- A major change from a one-particle bulk simulator to a many-particle device simulator.
- Improved time step algorithm.
- More accurate integration of the equations of motion.
- A self-consistent coupling between these equations and Poisson's equation.

The system we use in most of this study is shown in Fig. 1. It is a Si Schottky diode with doping $N_D = 10^{17} \text{ cm}^{-3}$ and a conduction band bending of 0.025–0.1 V. The contact on the l.h.s. is ohmic via a neutral region. The r.h.s. is treated as an absorbing boundary. The neutral region is kept neutral by uniform injection of electrons with a Fermi-Dirac distribution in energy. The simulations start with an initial solution close to steady state with a duration of about 6 ps. We have tried to keep the number of simulated particles small, about 40 particles per mesh cell have been used. If we should use a mesh in two dimensions with e.g. 30×30 cells the total number of particles would be about 30 000. Poisson's equation is solved with an interval of 5 fs using a 4–5th order Runge-Kutta method giving a 4th order truncation error [10]. Tunnelling across the Schottky barrier has been included using the WKB method [11].

3. Potential fluctuations and the lookup table approach

In the undepleted region of the Schottky diode the field is very low, and depends strongly on fluctuations in the charge

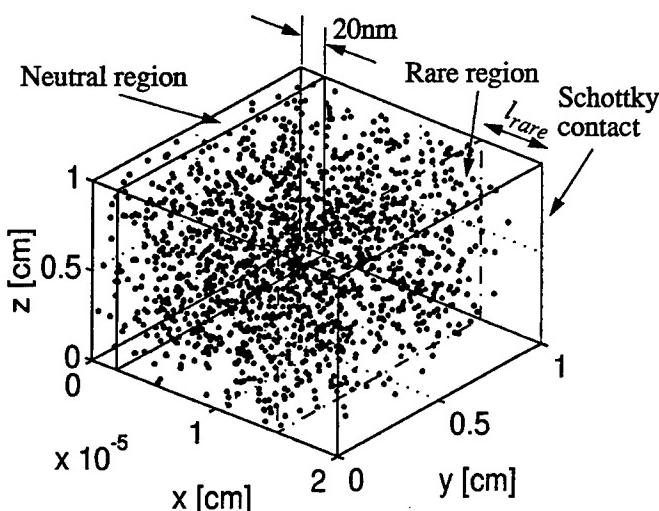


Fig. 1. Schottky diode structure used in the simulation. $l_{\text{rare}} = 50 \text{ nm}$. The electron population shown is for a band bending of 0.1 V.

density. The assignment of charge to the mesh points is done very often since Poisson's equation is solved frequently in order to get a consistent simulation in time. Some of the assignment functions being used in MC simulations are shown in Fig. 2, in addition we also show some more complex ones. The nearest grid point (NGP) and cloud in cell (CIC) have the advantage that they are easy to implement and fast to perform, but if one studies the effect of the charge assignment on the potential it is clear that they can cause problems. In Fig 3, we show the results from a calculation where 1000 electrons have been distributed randomly 1000 times over an interval of $1 \mu\text{m}$ in space. Poisson's equation was solved with the potential equal to zero on each boundary and the maximum voltage deviation was recorded. The size of the charge cloud is important as can be seen if we compare TSC and HSC 4-points with the corresponding 2-point schemes. Clearly the triangular shaped cloud (TSC) and the Hamming shaped cloud (HSC) give less potential fluctuations than NGP and CIC. The Hamming shaped cloud can be written as:

$$w(x) = \begin{cases} \frac{1}{\Omega} \left(0.54 - 0.46 \cos \left(2\pi \frac{x}{H} \right) \right), & -\frac{H}{2} < x \leq \frac{H}{2}, \\ 0, & \text{elsewhere} \end{cases} \quad (1)$$

where Ω is a normalisation constant and H is the size of the cloud.

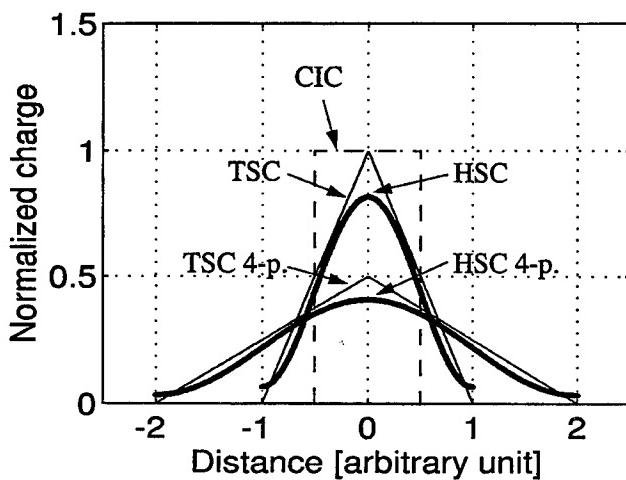


Fig. 2. Assignment functions used in this paper.

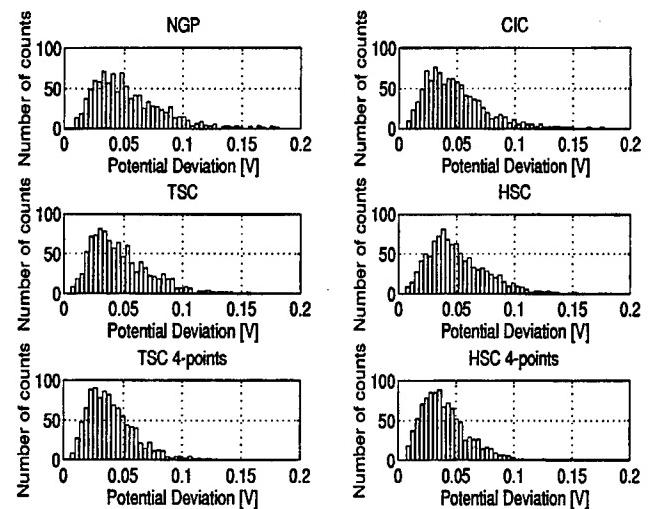


Fig. 3. Histograms over potential fluctuation using different assignments functions.

In principle one could use a scheme with different assignment functions in different spatial regions. In [7] the concept of self force is discussed at length. The theory shows that self force is introduced for non uniform mesh cells and when different assignment schemes are used in different regions. However, in spite of this, state of the art simulators used non uniform mesh assignment [3]. In Fig. 4, we compare the electric fields for uniform charge assignment and non uniform charge assignment, using different assignment functions in each half of the interval. The input data to the charge assignment functions is identical. For the cases that we have tested we can not find any significant problems. However, each case has to be judged separately.

For more complex assignment schemes a lookup table approach is convenient since the CPU time will not depend on the particular assignment function. It is desirable to separate this part from the "MC kernel" of the program. Some advantages of this approach are:

- (1) Each charge assignment can be performed using only one floating point operation and a number of integer operations.

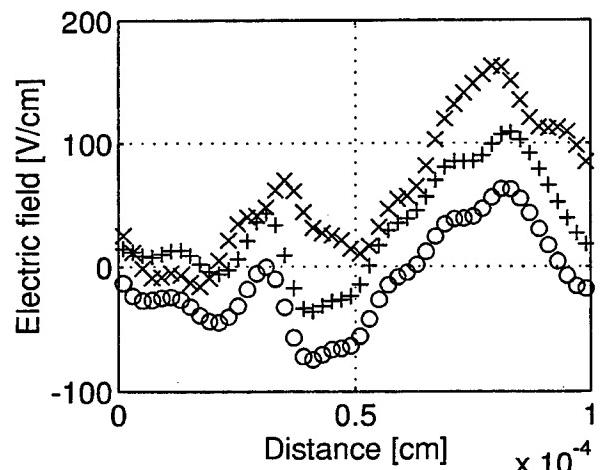


Fig. 4. Mean electric field in a semiconductor interval ($N_b = 10^{16} \text{ cm}^{-3}$) with zero Volt at the boundaries. (x) CIC, (○) TSC 4-p. (+) CIC in the left half and TSC 4-p. in the right half. The self force can be regarded as much smaller than the statistical noise in the electric field.

- (2) The CPU time required for a charge assignment is independent of the form of the cloud as well as of the number of boundaries.
- (3) Several different assignment functions can be used without extra requirements in CPU time.

The charge assignment is based on a technique similar to that used in [12]. In our approach a lookup table is divided into two data arrays, one containing the charge values to be assigned and another containing the relative index to the mesh points used in the assignment. We refer to these arrays as the assignment data array and the control frame array. We will describe one-dimensional charge assignment but extension to two dimensions is straightforward. The entire device is divided into a very fine particle position mesh (PPM). An absolute index to the nearest mesh point in the PPM can be calculated as

$$\text{ippm} = \text{int}(r/\text{ppmsize}) + 1 \quad (2)$$

where *int* means integer part, *r* is the position coordinate and *ppmsize* is the mesh size of the PPM.

The field mesh (FM) used when Poisson's equation is solved contains much fewer points than the PPM. The index to nearest FM point can be calculated as

$$\text{iabs} = \text{div}(\text{ippm}, \text{ppmnum}) + 1 \quad (3)$$

where *ppmnum* is the number of PPM intervals between two points in the field mesh and *div* means integer division.

The local index to nearest field mesh point is

$$\text{irel} = \text{mod}(\text{ippm}, \text{ppmnum}) + 1 \quad (4)$$

where *mod* means modulus.

The absolute index *iabs* is used as a key to a control table with indices to data entries in the control frame array. The relative index *irel* is used as an index in the assignment data array. Charge values are read from the assignment data array and scaled by the superparticle weight factor before assignment. Node indices are calculated as *iabs* plus the relative indices read from the control frame. A pointer diagram describing the charge assignment can be seen in Fig. 5.

In a two-dimensional mesh using an ordinary CIC assignment scheme the assignment of *N* particles consists of *N* floating point operations and 2^*N integer operations plus 4^*N data assignments. The storage requirements for the entire lookup table depends on the number of PPM points within one field mesh cell and the cloud size. A CIC method need 4 data fields in each charge assignment while a 4-point

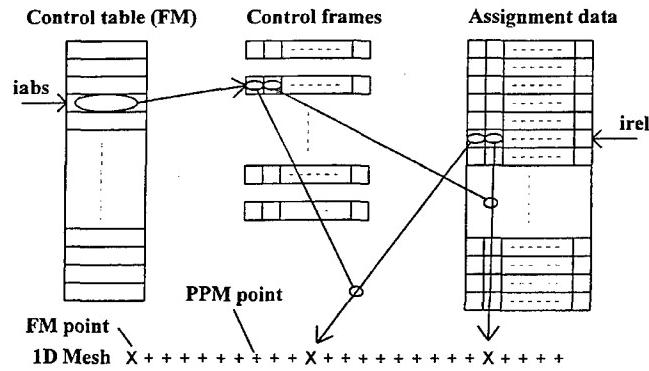


Fig. 5. Schematic diagram over a charge to mesh assignment in one dimension.

TSC method needs 16 data fields. The size of the control table depends on the number of field mesh points in the device. The number of the control frames needed depends strongly on the number of boundaries within the device. A two-dimensional Schottky diode with a 4-point TSC method needs 25 control frames. Both the control table and the control frame are integer arrays with small storage requirements. The large part is the data array which is a floating point array.

The implementation of the method described reduces the time needed to do the charge assignment to about one third of the time for a conventional scheme.

4. Monte Carlo results

We have performed MC simulations of a Schottky diode described in Section 2. The potential distribution is shown in Fig. 6. A comparison between the values of the current density for different assignment functions is shown in Fig. 7. As we can see, the magnitude of the current does not depend on the assignment function. However, the noise in

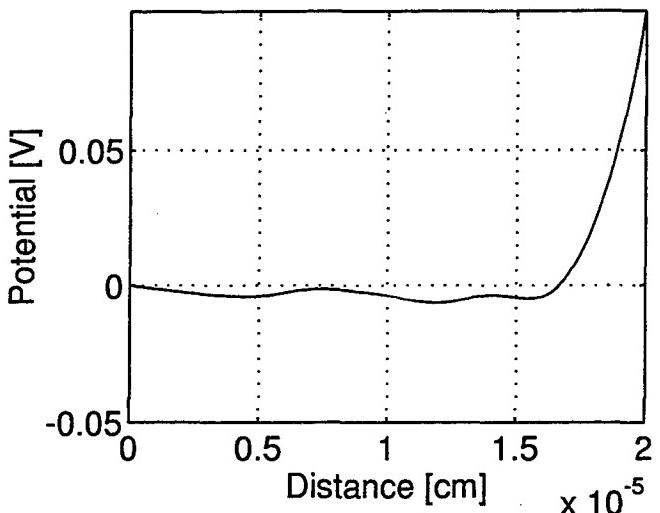


Fig. 6. Potential in the diode for a TSC 4-p. cloud. In the low field region we can see potential fluctuation due to noise in the charge distribution.

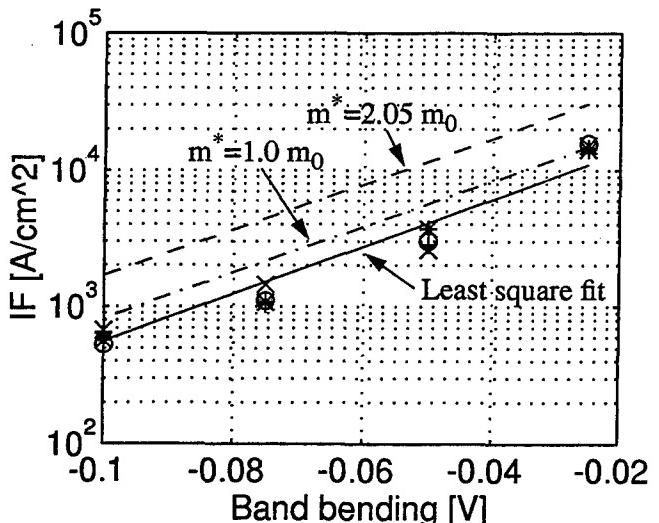


Fig. 7. Current density obtained for different assignments functions. (*) NGP, (x) CIC, (+) TSC 4-p., (O) HSC 4-p. Dashed lines – current density obtained from the thermionic emission theory.

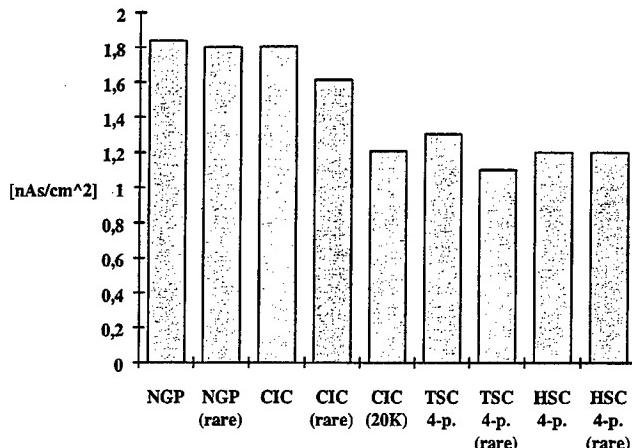


Fig. 8. Standard deviation of the absorbed charge at the Schottky contact for different assignment functions and superparticle densities. 2000 superparticles were used except CIC 20K, which is a simulation using the CIC assignment function with 20 000 superparticles.

the charge absorbed at the Schottky boundary shows a strong dependency on the assignment function as can be seen in Fig. 8. In Fig. 7 the MC result is compared with the thermionic emission (t.e.) theory. The MC simulation gives lower current levels than the t.e. theory. A possible explanation is that the algorithm for interpolating the scattering rate at low energies is not sufficiently accurate. In [13] a Si Schottky diode has also been simulated with an Ensemble Monte Carlo method but with a simplified band structure. The current levels shown there are also lower than the values predicted by t.e. theory.

5. Rare regions

Possible ways of improving the statistics in high energy regions rarely visited by charge carriers were introduced and applied to bulk material in [14]. The idea of "splitting" the particles when they visit certain "rare" regions has been extended to devices as well as to regions in real space [1, 3]. In [15, 16] an extension of the method was introduced where particles were split in some regions and put together in others so that the total amount of mobile charge in the device did not change with time. We have followed the same path and implemented a rare region near the Schottky contact. When an electron reaches the rare region its weight factor (WF) is compared with the weight factor of the rare region. If WF for the electron is larger than WF for the rare region the electron is split into a subset of rare electrons with WF equal to WF for the rare region. A rare electron in the normal region has a WF less than WF for the normal region and tries to find other rare electrons in the same situation. If such an electron exists one of the electrons absorbs the other and the WF is updated. The critical part of this scheme is the absorption of electrons in the normal region. In a simulation with one dimensional boundaries the boundary of the rare region is a single point and the region where the absorption process takes place is localized in space. As a result the perturbation of the field is very small and can be neglected. The magnitude of the current is not affected by the introduction of a rare region. However, the noise in the absorbed charge at the Schottky barrier decreases (Fig. 8). We have only used one rare region but in principle it is possible to use an arbitrary number. One pos-

sible solution is to assign a WF for each mesh interval in such a way that the number of carriers is constant throughout the whole device. In the Schottky diode we did not find any significant reason to go beyond the single rare region approach. However, this method could make it possible to study transport over larger barriers.

6. Conclusions

Monte Carlo simulation of a Si Schottky diode has been presented. Different methods of statistical enhancements have been discussed. It has been shown that more complex assignment functions is an effective way of reducing the number of particles needed in order to reduce the noise in MC quantities. Simulation results shows that the 4-point Hamming cloud using 2000 particles gives about the same noise as a conventional CIC with 20 000 particles (Fig. 8). A charge assignment method built on a table lookup approach has been presented that can handle complex assignments schemes without suffering from large CPU time requirements. Our simulations show that using high order assignment functions it is possible to simulate forward biased diodes in 2D structures like the PBT without an excessively large number of superparticles. The current density obtained for the Schottky diode is smaller than what is predicted from thermionic emission theory. An explanation of this effect needs further investigation. Statistical enhancements using a rare region has been implemented and it has been shown that this is one way of increasing the resolution in MC quantities. However, we have found that the noise in the absorbed charge at a Schottky contact is more sensitive to the choice of assignment function in the entire device than on the number of particles near the absorbing boundary.

Acknowledgements

The authors would like to thank Prof. Karl Hess and Dr Carl Wordelman, University of Illinois, for providing the source code and input data to the Monte Carlo program in [9]. Financial support from the Mid-Sweden University is gratefully acknowledged.

References

1. Jacoboni, C. and Lugli, P., "The Monte Carlo Method for Semiconductor Device Simulation" (Springer-Verlag, Wien 1989).
2. Hess, K. (Editor), "Monte Carlo Device Simulation: Full Band and Beyond" (Kluwer Academic Publishers 1991).
3. Laux, S. E. and Fischetti, M. V., "Monte Carlo Device Simulation: Full band and Beyond" (Edited by K. Hess) (Kluwer Academic Publishers 1991), ch. 1, p. 1.
4. Kunikiyo, T. et al., J. Appl. Phys. 297 (1994), vol. 75, no. 1.
5. Cheng, D. Y., Hwang, C. G. and Dutton, R. W., IEEE Trans. on Computer-Aided Design 7, 1017 (1988).
6. Kosina, H. and Selberherr, S., IEEE Trans. On Computer-Aided Design 13, 201 (1994).
7. Hockney, R. W. and Eastwood, J. W., "Computer Simulation Using Particles" (Adam Hilger, Bristol 1988).
8. Birdsall, C. K. and Fuss, D., Journal of Computational Physics 3, 494 (1969).
9. Shichijo, H., Tang, J. Y., Bude, J. and Yoder, D., "Monte Carlo Device Simulation: Full band and Beyond", (Edited by K. Hess) (Kluwer Academic Publishers 1991), ch. 10, p. 285.
10. Forsythe, G. E., Malcolm, M. A. and Moler, C. B., "Computer methods for mathematical computations" (Prentice Hall 1977).
11. Bransden, B. H. and Joachain, C. J., "Introduction to Quantum Mechanics" (Longman Scientific & Technical 1989), ch. 8, p. 397.

12. Kosina, H. and Selberherr, S., "Simulation of Semiconductors Devices and Processes" (Edited by D. Aemmer and W. Fichtner) (Hartung-Gorre, Konstanz 1991), p. 251.
13. Ravaioli, U., Lugli, P., Osman, M. A. and Ferry, D. K., IEEE Trans. on Elec. Dev. **ED-32**, 2097 (1985).
14. Phillips, A. Jr., and Price, P. J., Applied Physics Letters, **30**, 528 (1977).
15. Sangiorgi, E., Ricco, B. and Venturi, F., IEEE Trans. on Computer-Aided Design **7**, 259 (1988).
16. Venturi, F., Smith, R. K., Sangiorgi, E. C., Pinto, M. R. and Ricco, B., IEEE Trans. on Computer-Aided Design **8**, 360 (1989).

Cobalt Disilicide (CoSi_2) Schottky Contacts to 6H-SiC

N. Lundberg and M. Östling*

Royal Institute of Technology, Department of Electronics, P.O. Box E229, S-164 40 Kista-Stockholm, Sweden

Received June 2, 1994; accepted in revised form June 27, 1994

Abstract

Schottky contacts using CoSi_2 to both n- and p-type 6H-SiC were fabricated. The contacts revealed good rectifying characteristics after annealing at 700 °C. Low leakage currents and exponentially increasing currents over at least 5 decades were obtained in the forward bias mode. C-V- and I-V-measurements were used to establish the Schottky barrier heights for CoSi_2 to 6H-SiC, 1.05 ± 0.05 eV and 1.90 ± 0.05 eV for n- and p-type respectively. Further annealing at 900 °C changed the Schottky barrier heights on both n- and p-type significantly.

1. Introduction

Silicon carbide has many material properties that make it suitable for use in high temperature devices. An important requirement is temperature stable electrical contacts that withstand temperatures in the excess of 500 °C. Refractory metal silicides have been suggested as contact material because of their high temperature stability, low resistivity and compatibility with existing IC-technology [1, 2]. Cobalt disilicide (CoSi_2) fulfills most of the criteria mentioned above. Cobalt is an interesting choice also from a more fundamental point of view because in the binary material system Co-C, no cobalt carbides are reported [3]. Consequently when Co reacts with the Si in the SiC substrate, residual carbon will be left. In metal-semiconductor junctions the Schottky barrier height (SBH) describes the electrical characteristics of the contact.

In this study we investigate CoSi_2 contacts formed on SiC after deposition of Si and Co thin films on 6H-SiC. The relative thickness of the deposited films resulted in a metal-rich CoSi_2 after annealing. The barrier height for the semiconducting (SiC)/metal silicide interface (mostly CoSi_2 and a fraction of CoSi) was investigated. Annealing at high temperature could lead to a substrate reaction; silicon from the substrate reacts with CoSi and forms CoSi_2 and free carbon precipitates. Earlier studies of a similar reaction [4, 5] have revealed that residual carbon forms precipitates 0.5 μm in diameter, homogeneously distributed throughout the silicide with a distinct accumulation at the surface. Therefore, two Co/Si films with different compositions were deposited on SiC to find out if the carbon precipitates influence the electrical behaviour of the Schottky contacts. One with (Co-Si, 36–64/SiC), and one with (Co-Si, 38–62/SiC).

2. Experimental

Highly doped single crystal substrates of both n-type (nitrogen-donors) and p-type (aluminium-acceptors) 6H-SiC,

containing 5 μm thick low doped epitaxial films, from Cree Research, Inc. were used. The cleaning sequence prior to deposition consisted of degreasing in organic solvents, 2 min in a $\text{H}_2\text{SO}_4 : \text{H}_2\text{O}_2$ (3 : 1) solution, rinsing in water for 5 min and blown dry to N_2 . Cobalt was evaporated from an e-gun heated source at a rate of 0.2 nm/s to a thickness of 44 nm. Silicon was then evaporated at a rate of 0.5 nm/s to a thickness of 127 nm. The chamber was then filled with nitrogen and quickly opened, and some samples were removed. After a pump down a second evaporation of Si was performed at the same rate as previously deposition until the thickness of the silicon layer had increased to 145 nm. This procedure was done in order to create two depositions with the same amount of Co but with two distinct Si amounts. The base pressure during all evaporation was $6 \cdot 10^{-6}$ Pa. A standard "lift-off" procedure was used to obtain circular contact areas 86 μm and 173 μm in diameter. All annealings were performed in a vacuum furnace at a base pressure of $2 \cdot 10^{-5}$ Pa at 500 °C, 700 °C and 900 °C, for 5 h, 2 h and 2 h, respectively. X-ray diffraction (XRD) measurements were performed using a Cu-K α_1 radiation source to identify the phases formed. Determination of the compound composition and the depth profile was done by Rutherford Back-scattering Spectrometry (RBS), utilising 2.4 MeV ${}^4\text{He}^+$. Inspection of the surface morphology was carried out by 15 kV Scanning Electron Microscopy (SEM). All electrical measurements were performed with InGa ohmic contacts on the backside. The InGa was carefully removed in an ultrasonic ethanol bath before every subsequent annealing. A HP-4145A parameter analyser was used to record the I-V-measurements and C-V-measurements were performed using a HP-4280A 1 MHz C-meter interfaced to a HP-486/33N computer. All capacitance-voltage values correlated to a straight line in the $1/C^2$ vs. applied voltage diagrams, and the regression coefficient was always at least 0.9999. I-V-measurements were performed at temperatures ranging from 280 K to 475 K.

3. Results

3.1. Si/Co/SiC solid state reactions

After annealing at 500 °C, XRD showed that peaks identified as CoSi ($d = 3.13$ Å) and CoSi_2 ($d = 1.90$ Å, 1.62 Å and 1.09 Å) developed. No Co or Co_2Si peaks were detected. Further annealing at 700 °C and 900 °C did not change the spectra significantly. Some of the peaks from CoSi decreased in intensity whereas peaks from CoSi_2 grew stronger.

Characterisation of the two depositions was done by RBS. Both the Co and the Si signals were clearly separated from the substrate signal and hence the layer thicknesses could be determined with good accuracy. The two depositions resulted in Co-Si (36–64) and Co-Si (38–62) films.

* Present address: Stanford University, CIS, Stanford, CA 94305.
e-mail address: nils@ele.kth.se

Heat treatment at 500 °C leads to formation of CoSi_2 . Examination of the Co-Si (38–62) film reveals the Co concentration to decrease from 38 atomic percent at the SiC/ CoSi_2 interface to 34 atomic percent at the surface. This observation leads to two conclusions: CoSi is present in the silicide film and the intermixing of Co and Si is not complete. Annealing at 700 °C does not further distribute the two elements. Annealing at 900 °C completed the intermixing of Co and Si in the silicide. Figure 1 displays two plateaus for the Si signals revealing a difference in average stoichiometry for the contacts which indicates no reaction with the substrate.

SEM micrographs of the surface after annealing at 500 °C and 700 °C displayed a homogeneous surface. However, after annealing at 900 °C precipitates 0.3 µm in diameter occur at the surface. The density of precipitates per unit area was larger on the Co-Si (38–62) film compared to the Co-Si (36–64) film. The appearance of the precipitates could be evidence of a slight substrate reaction.

The result of the different materials analysis can be summarised as follows. Annealing at 500 °C for 5 h formed CoSi and CoSi_2 [6]. (All Co was consumed and no substrate reaction occurred.) After annealing at 700 °C no evidence of substrate reaction could be found. Slight substrate reaction might have occurred during the 900 °C anneal but remaining traces of CoSi in the film reveal a non-complete CoSi_2 formation.

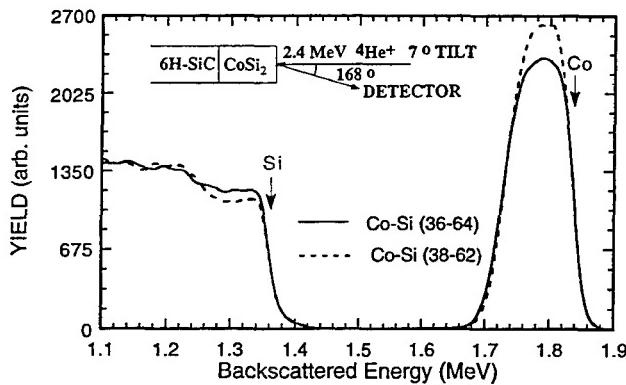


Fig. 1. RBS-spectra of the Schottky contacts revealing the difference in composition between the two depositions after annealing at 900 °C. The lower Co signal and the higher Si signal belongs to the Co-Si (36–64) film. The arrows correspond to the surface positions of the elements and the inset depicts the sample configuration during analysis.

Table I. Results of Schottky-barrier height ϕ_B measurements of n-type Co-Si (36–64)/6H-SiC contacts by capacitance-voltage ($C-V$), current-voltage ($I-V$) and temperature activated- $I-V$ (act.) methods as a function of annealing temperature; also included are measurements of the reverse leakage current density at 5 V reverse bias as a function of annealing temperature

Annealing temperature (°C)	n-type				
	$\phi_{B,C-V}$ (eV)	$\phi_{B,I-V}(\text{temp})$ (eV)	η	$\phi_{B,\text{act}}(\text{temp-interval})$ (eV)	J_{rev} (A/cm ²)
500	1.05	1.01 (300 K)	1.15	0.96 (300 K–350 K)	$3.0 \cdot 10^{-7}$
700	1.05	0.82 (300 K)	1.16	0.75 (300 K–350 K)	$4.2 \cdot 10^{-5}$
		0.91 (450 K)	1.06	0.86 (350 K–475 K)	
900	0.75	0.52 (300 K)	1.25	0.41 (300 K–475 K)	$1.7 \cdot 10^0$
		— (450 K)	2.2		

3.2. Electrical characteristics of n-type rectifying contacts

In a Schottky rectifier fabricated on a semiconductor with low carrier concentration, current components resulting from field emission and tunnelling are of minor importance and can be neglected. The only significant current contribution is thermionic emission where electrons transport over the potential barrier into the metal. Thus, the current density across the interface under the application of bias voltage is given by [7]

$$J = J_0(e^{qV/\eta kT} - 1) \quad (1)$$

where

$$J_0 = BT^2 e^{-\phi_B/kT} \quad (2)$$

$B = 71.8 \text{ A/cm}^2/\text{K}^2$ is the effective Richardson constant for 6H-SiC, ϕ_B is the Schottky-barrier height in SiC, and η is the ideality factor. The log J vs. forward voltage characteristics for the two different film compositions exhibited some changes in the diode characteristics after the three annealings. After 500 °C, distinct difference between the two compositions appears in the characteristics, which decreases after the two following annealings. After the 500 °C annealing the SBH for the Co-Si (36–64) contact has its highest value, 1.01 eV, and then decreases after annealing at higher temperatures as can be seen in Table I. It reaches its lowest value, 0.52 eV, after annealing at 900 °C. The ideality factor η remains almost constant at 1.2 after the heat treatments although slightly lower slope of the curves annealed at 900 °C reveals an increased ideality factor. The lower slope seen after the highest annealing temperature can be interpreted as an increased recombination current, in the depletion region of the contact. Forward characteristics of the Co-Si (36–64) contact after annealing at 500 °C as a function of temperature can be seen in Fig. 2. Good behaviour over at least 6 decades are demonstrated with the ideality factor always lower than 1.2 in the linear region. After annealing at 700 °C, the ideality factor decreases and the SBH increases at higher operating temperatures. At 475 K the contact displays almost ideal characteristics with the ideality factor as low as 1.06.

At 5 V reverse bias the current density for the Co-Si (36–64) contact was $2.1 \cdot 10^{-7} \text{ A/cm}^2$ after annealing at 500 °C. Annealing at 700 °C and 900 °C increased it to $4.5 \cdot 10^{-5} \text{ A/cm}^2$ and 1.7 A/cm^2 , respectively. Obviously the leakage current varies inversely with the SBH. Further reverse biasing to 100 V for the contact annealed at 500 °C only

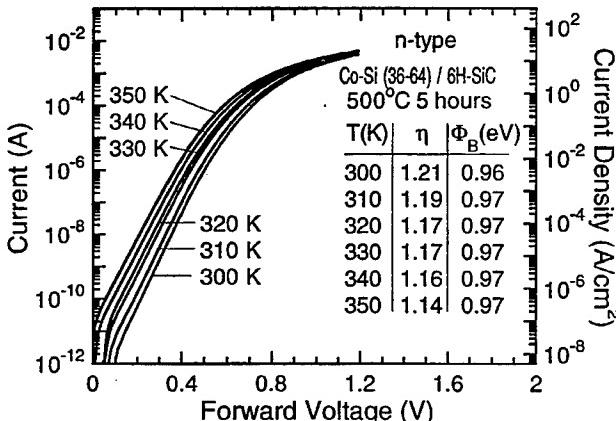


Fig. 2. Forward current-voltage characteristics for a n-type Co-Si (36-64)/6H-SiC contact, annealed at 500 °C.

increased the current density to $5 \cdot 10^{-5} \text{ A/cm}^2$. Irreversible breakdown did not occur for the contacts under a reverse bias of 100 V.

Log J/T^2 vs. $1/T$ plots were used to investigate the contribution of recombination current. Annealing at 500 °C: at temperatures ranging from 300 K to 350 K the activation energy was 0.96 eV.

Annealing at 700 °C: measurement from 350 K to 475 K yielded two activation energies, E_A 's, equal to 0.86 eV and 0.75 eV at the high and low temperature end of the interval, respectively. The SBH at 475 K was 0.91 eV and the ideality factor was 1.06. A linear fit of the $\phi_B(T)$ values obtained could be extrapolated with good regression to the SBH at zero temperature.

Annealing at 900 °C: only one activation energy ($E_A = 0.41$ eV) was found. The SBH at 300 was 0.61 eV. The ideality factor increased from 1.26 to an extracted value of 2.2 when the temperature was increased from 300 K to 450 K. These results indicate a large increase of recombination after annealing at 900 °C. Negligible difference in the $I-V$ -characteristics with regards to the Co-Si (36-64) contact were obtained for the Co-Si (38-62) contact.

Figure 3 displays $C-V$ -data for the n-type contacts. The highest SBH 1.06 eV, corresponds to annealing at 500 °C. Further annealing at 700 °C only has a small effect on the SBH. Annealing at 900 °C lowers the barrier to 0.75 eV. The

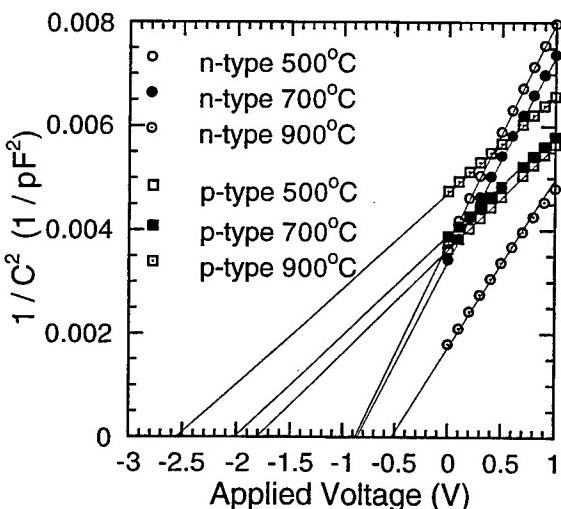


Fig. 3. $C-V$ -data of the Co-Si (36-64)/6H-SiC contacts after the different annealings.

extracted dopant concentration increases from $5 \cdot 10^{16}$ to $7 \cdot 10^{16} \text{ cm}^{-3}$ after annealing at 500 °C and 900 °C respectively. The SBH was about 0.06 eV lower for the Co-Si (38-62) film over the investigated interval.

3.3. Electrical characteristics of p-type rectifying contacts

Electrical characterisation after the 500 °C anneal revealed the presence of an insulating layer at the SiC/silicide interface. However, after annealing at 700 °C for 5 min the contacts displayed rectifying characteristics. Materials analysis did not reveal any differences before and after this latter heat treatment. These observations suggest a break up process of the insulating layer during the 5 min annealing at 700 °C.

The forward current-voltage characteristics of the p-type Co-Si (36-64) contacts annealed at 500 °C and 5 min at 700 °C are shown in Fig. 4. From the linear portion of the curve an ideality factor of 1.38 and a SBH of 1.41 eV was obtained at 300 K. At 350 K the ideality factor improved to 1.31. The large differences in current density at 1.2 V can be attributed to incomplete ionisation of the p-type dopants. Annealing at 900 °C decreased the linear region to less than three decades and the SBH to 1.15 eV as can be seen in Table II. As for the n-type contacts the reverse leakage current of the p-type contacts was proportional to the SBH. No irreversible breakdown occurred before 100 V and the current density was $5.3 \cdot 10^{-5} \text{ A/cm}^2$ when the contacts were reverse biased to 100 V after the 500 °C annealing.

Temperature activation measurements for p-type Co-Si (36-64) contacts annealed at 500 °C were done in the 280 K to 350 K temperature range. An E_A of 1.04 eV was obtained and the $\phi_B(T)$ values extracted from the log J vs. V_{for} plots showed good correlation to a linear interpolation. The zero temperature SBH was 1.00 eV. Annealing at 700 °C increased the recombination current. Measurement at temperatures ranging between 300 K to 475 K, yielded two E_A 's, one at 1.16 eV corresponding to the thermionic emission current and a second one below 400 K equal to 0.64 eV. Annealing at 900 °C: activation plots from temperatures ranging between 300 K and 475 K, revealed E_A equal to 0.77 eV. Starting from values around 1.4, the extracted ideality factor increased to 2.5 when the temperature was increased from 300 K to 450 K.

The results from the $C-V$ -measurements of p-type Co-Si (36-64) contacts are displayed in Table II and in Fig. 3.

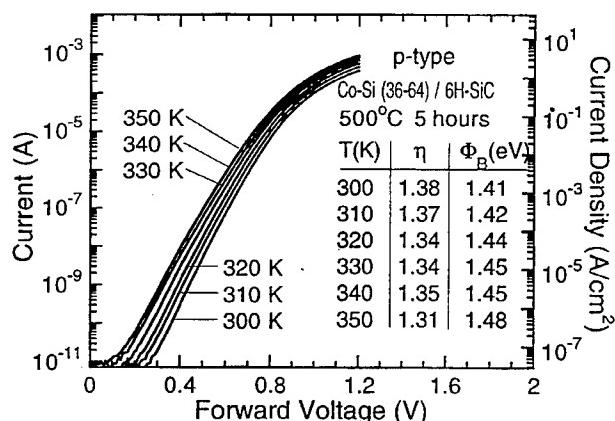


Fig. 4. Forward current-voltage characteristics for a p-type Co-Si (36-64)/6H-SiC contact, annealed at 500 °C.

Table II. Results of Schottky-barrier height ϕ_B measurements of p-type Co-Si (36–64)/6H-SiC contacts by capacitance-voltage (C-V), current-voltage (I-V) and temperature activated I-V (act.) methods as a function of annealing temperature; also included are measurements of the reverse leakage current density at 5 V reverse bias as a function of annealing temperature

p-type					
Annealing temperature (°C)	$\phi_{B\text{C-V}}$ (eV)	$\phi_{B\text{I-V}}(\text{temp})$ (eV)	η	$\phi_{B\text{act}}(\text{temp-interval})$ (eV)	J_{rev} (A/cm ²)
500	1.90	1.41 (300 K)	1.38	1.04 (300 K–350 K)	$6.3 \cdot 10^{-7}$
		1.48 (350 K)	1.31		
700	2.10	1.30 (300 K)	1.61	0.61 (300 K–350 K)	$8.1 \cdot 10^{-6}$
		1.54 (450 K)	1.41	1.16 (350 K–475 K)	
900	2.69	1.15 (300 K)	1.33	0.61 (300 K–475 K)	$2.4 \cdot 10^{-5}$
		— (450 K)	2.5		

Dopant concentrations of $8 \cdot 10^{16} \text{ cm}^{-3}$ and $1 \cdot 10^{17} \text{ cm}^{-3}$ were extracted after heat treatments at 500 °C and 900 °C. The SBH increased after every heat treatment. After annealing to 900 °C the extracted value for the SBH was 2.69 eV. Obviously the C-V-method over-estimate the SBH when the slight substrate reaction has occurred. The Co-Si (38–62) contact displayed similar results for all electrical measurements.

4. Discussion

4.1. Solid state reactions

No detectable reaction between the substrate and the silicide took place after annealing at 700 °C. This observation agrees with an earlier report of an increased reaction temperature for silicidation on SiC compared to Si [5].

4.2. Schottky barrier height and recombination for n-type contacts

The SBH obtained from C-V- and I-V-measurements after the 500 °C annealing are in good agreement. The result obtained from temperature activation corresponds to the SBH at zero temperature. As the operating temperature is increased the recombination contribution decreases yielding a higher value of the SBH. After the 700 °C anneal the recombination current increased. Annealing at 900 °C changed the rectifying characteristics substantially, the SBH was lowered to 0.30 eV. This observed lowering of the SBH after annealing, is in agreement with findings by Porter [8]; however, for the Pt/SiC system, the barrier height is reported to increase after substrate reaction [9]. These results are somewhat surprising since the two materials systems, Co-SiC and Pt-SiC, resemble each other with respect to the absence of metal-carbides.

The absence of irreversible breakdowns when the contacts were reverse biased to 100 V is attributed to the lack of edge termination [10]. On the other hand, Bhatnagar *et al.* [11] claims that the ideal value for the breakdown voltage should be a factor 1.4 higher than the value obtained without surface passivation and edge termination. Hence it can be expected that no breakdown should occur before 140 V.

4.3. Schottky barrier height and recombination for p-type contacts

If the values of the SBH obtained by both C-V- and I-V-methods are put together as in Table II they display a

complex behaviour. Annealing below the threshold temperature for substrate reaction, yield SBHs and ideality factors around 1.4. At higher operating temperatures the recombination current decreases. Results obtained by C-V-measurement yield higher values of the SBH, 1.90 eV and 2.10 eV after annealing at 500 °C and 700 °C, respectively. A large amount of the discrepancy between the SBHs can be attributed to the high value of the ideality factor. Annealing at 900 °C leads to higher ideality factors. η has a value of 1.58 at 300 K and reaches a value of 2.5 at 450 K. These unusual high values of η for both the n- and p-type contacts at elevated temperatures could be explained by traps created at the interface by the substrate reaction during the 900 °C annealing. As the operating temperature increases more carriers are generated that can recombine thereby yielding a higher value of η . C-V- and I-V-measurements reveal a SBH of 2.69 eV and 1.34 eV, respectively. The overestimation of the SBH obtained by C-V-measurements can be explained by an additional in-series capacitance in the contact. For instance, the Al dopants at the SiC surface can react with CoSi₂ and form a Co₂Al₉ alloy [12] and leave a thin intrinsic SiC layer underneath that introduces additional series capacitance. Additional research is needed to confirm these speculations.

The results after the annealings suggest that the slight substrate reaction affects the value of the SBH. At temperatures below 900 °C where no evidence of reaction with the substrate is found, a value of 1.05 ± 0.05 eV and 1.90 ± 0.05 eV can be reported for the SBH of CoSi₂ to n- and p-type, respectively. Small islands of CoSi in the film could affect the SBH obtained if $\text{SBH}_{(\text{CoSi}_2)}$ was lower than $\text{SBH}_{(\text{CoSi})}$, but that is not the case since the SBH to n-type SiC is lowered when Si is supplied to the silicide after the slight substrate reaction. The sum of the SBHs on n- and p-type add up to 2.95 ± 0.1 eV which is close to the energy bandgap of 6H-SiC, $E_g(6\text{H-SiC}) = 2.86$ eV, at 300 K [13]. Annealing at 900 °C changes the obtained values of the SBH significantly. On n-type the barrier is lowered to 0.75 eV but increased on p-type to 2.7 eV. This rather large change of the SBH can mainly be attributed to the slight substrate reaction although other mechanisms are not ruled out for changing the electrical properties.

5. Conclusions

- Cobalt disilicide Schottky contacts to both n- and p-type 6H-SiC have been investigated. The contacts display stable

electrical response after annealing at 700 °C featuring good rectifying I - V -characteristics with low ideality factors over at least 5 decades and a reverse leakage current density of 10^{-5} to 10^{-7} A/cm^2 at a reverse bias of 5 V. No distinct breakdowns took place even at 100 V reverse bias. The Schottky barrier height (SBH) of CoSi_2 to 6H-SiC was $1.05 \pm 0.05 \text{ eV}$ and $1.90 \pm 0.05 \text{ eV}$ for n- and p-type, respectively.

2. Annealing at 900 °C increased the recombination current and changed the obtained SBH significantly.

3. A large discrepancy between SBHs was obtained by I - V - and C - V -measurement for the p-type contact after annealing at 900 °C. The C - V -method overestimates the SBH.

References

1. Kelner, G., Proc. of the 15th Nordic Semiconductor Meeting (Edited by S. Fransilla and R. Paananen) (Hämeenlinna, Finland 1992), p. 5.
2. Chaudhry, M. I., Berry, W. B. and Zeller, M. V., Mat. Res. Soc. Symp. Proc. **162**, 507 (1990).
3. Massalski, T. B., Murray, J. L., Bennett, L. H. and Baker, H. (Eds.), "Binary Alloy Phase Diagrams" (American Society for Metals, Metal Park, Ohio 1986), vol. 1, p. 556.
4. Lundberg, N. and Östling, M., Appl. Surf. Sci. **63**, 3069 (1993).
5. Lundberg, N., Zetterling, C.-M. and Östling, M., Appl. Surf. Sci. **73**, 316 (1993).
6. d'Heurle, F. M. and Petersson, C. S., Thin Solid Films, **128**, 283 (1985).
7. Rhoderick, E. H. and Williams, R. H., "Metal Semiconductor Contacts" (Oxford University Press, Oxford 1988).
8. Porter, L. M. et al., Mat. Res. Soc. Symp. Proc. **282**, 472 (1993).
9. Papanicolaou, N. A., Christou, A. and Gipe, M. L., J. Appl. Phys., **65** 3526 (1989).
10. Spellman, L. M. et al., "Amorphous and Crystalline Silicon Carbide IV" (Edited by C. Y. Yang, M. M. Rahman and G. L. Harris) (Springer-Verlag, Berlin 1992), p. 417.
11. Bhatnagar, M. and Baliga, B. J., IEEE Trans. **40**, 645 (1993).
12. Van Gurp, G. J. et al., J. Appl. Phys. **50**, 6915 (1979).
13. Poerschke, R. and Madelung, O. (Eds.), "Data in Science and Technology" (Springer-Verlag, Berlin 1991).

The Pt/Si(111) Interface and the Properties of Thin Pt Layers on Si

P. Morgen, B. Jørgensen and J. Gordon

Fysisk Institut, Odense University, Campusvej 55, DK-5230 Odense M, Denmark

Received May 2, 1994; accepted June 15, 1994

Abstract

Thin Pt- and Pt-silicide films on Si are currently used in microelectronics as Schottky diodes with high values of the electrical barrier (~ 0.8 eV). Such films also have metallic or near metallic sheet conductivities and are suited for interconnects. The choice of Pt is indicated by the reaction between Pt and Si which seems easy to control, at low temperatures, enabling so-called self-alignment of the silicide pattern. This is somewhat unexpected, however, when considering the bulk thermodynamic phase diagram of these elements. From this, processing temperatures of over 830°C should be needed for reaching the eutectic temperature and an even higher temperature (1210°C) needed to reach the most stable silicide composition, PtSi. Many times, studies of the Pt/Si system under UHV conditions of processing and purity have been interpreted as forming silicide *at the Si surface*, even at room temperature. However, it was noted that *no bulk silicide* is grown from thin Pt deposits by heating in UHV. Normally, an inhomogeneous composition profile is found with the top surface being enriched in silicon. In the present studies Pt was deposited slowly on the Si(111) 7×7 surface. This was done to follow its reaction and in-depth distribution profile during deposition, and the formation of the Schottky barrier in this step. Subsequent annealing was then studied. Further experiments studied the sensitivity of annealing to impurities (C and O), and to oxidation. It was found that during deposition of Pt some Si atoms are retained in the top surface, bonding with or imbedding themselves in the Pt with a silicide-like appearance of the valence band, Si(2p) core levels, and the Si ($L_{2,3}$ VV) Auger spectrum (XAES). However the distribution of Pt and Si below the surface during deposition is indicative of an almost homogeneous Pt film, with some signs of a combined island and layer growth, past the first monolayer of Pt. The dislodging of Si atoms occurs during completion of the first monolayer, and the reaction starts only after this is completed. Impurities in a Pt film, present before annealing, were found to create a homogeneous, stable metallic phase. Oxidation of various Pt/Si structures differs from that of clean Si surfaces at intermediate temperatures, where evidence for the formation of an oxide with (weak) features of SiO_2 is obtained.

Introduction

In the continued development and refinement of materials and processes for microelectronics the search for a deeper understanding particularly of silicon oxidation and metallization reactions seem more important than ever to overcome some of the bottlenecks predicted for ultra large scale integration towards the year 2000 [1]. In fact it does not today seem justified to claim that these very important processes of the microelectronics technology are sufficiently well understood in basic terms. As a matter of fact there exists a kind of “pressure gap” in microelectronics between fundamental research, such as UHV based studies, and technological processing, as quoted for catalysis research. The advent of MBE methods seems to have brought with it, however, a closer match between basic and applied research, and maybe the next decade will see similar progress in oxidation and metallization research.

In the present project the properties of thin Pt films on Si(111) 7×7 surfaces are studied under UHV conditions. The aim was to characterize in more detail the reaction between Pt and Si, the interface and the growth mode of the film during deposition, including its in-depth composition. The effects of annealing with and without impurities were also studied, as was the possible catalytic activity of the metal or reacted compound towards oxidation. The band bending, and thus the Schottky barrier, can be followed in surface core level photoemission experiments with synchrotron radiation. These studies follow a number of previous reports [2, 3, 4, 5], in which the principal aim was to understand the relationship between the atomic and electronic structure of the interface and the Schottky barrier characteristics. Parallel studies looked for the best methods to form thick silicide layers and the possible epitaxial growth of PtSi on Si, such as rapid thermal annealing or the use of “forming” gases [6]. The failure of UHV based methods to produce homogeneous silicides were first noted in [2], and then further demonstrated in [4]. It was therefore uncertain, what Schottky barriers would result from UHV based processing.

Experimental details

The experiments were performed in UHV systems with electron spectrometers, Pt evaporators, gas inlet systems, excitation sources for photoemission, and sample cleaning and heating. Synchrotron radiation was available at MAX Lab in Lund, while all other experiments were done in Odense.

The Si(111) samples were P-doped n-type wafers with doping levels of 10^{17} cm^{-3} . They were cleaned by flashing after bake-out of the system, during which they were protected by a native oxide layer. For reference purposes Pt was also evaporated on the as received samples after bake-out to study if the growth of Pt differs from the case of a clean surface. Two different Pt evaporation methods were employed: At MAX Lab, a W wire supporting a thin Pt wire wound around it was heated, while in Odense an indirectly heated Ta plate supporting a Pt foil was used. The evaporation rate was monitored with a quartz crystal, elaborated into a double crystal set-up to compensate for thermal drift. Even with this system, however, radiation and charge pick-up turned out to be a problem, and only a short exposure period could be used to establish the rate, by subtracting the frequencies just before and shortly after exposure, with minimum thermal loading of the system. Further estimates of the rates were obtained from AES and XPS signals, and from the knowledge of the systematics. This

obviously gives a degree of uncertainty with respect to the absolute rates, but this is a problem common to all studies of this system.

Results

Deposition at room temperature

The growth of Pt on Si(111) 7×7 at a rate of deposition of 1 Å/min is demonstrated in Figs 1 and 2. Figure 1(a) shows the Si(2s), Si(2p) and Pt(4f) core lines and the background in that part of the XPS spectrum of the sample excited with Mg-K α radiation. Similarly, Fig. 1(b) shows the Pt(N,OO) and Si(L_{2,3}VV) Auger spectra, excited with the X-rays. In

Figs 2(a) and 2(b), the intensities of the signals are plotted as functions of the Pt dose received. Due to a difference in sampling depths at the different kinetic energies in these spectra, the two sets of intensities behave qualitatively different. The core line signals [Fig. 2(a)] indicate a nearly uniform growth of a Pt film on top of the Si surface, with a small amount of islanding [7]. They sample the entire Pt layer with only a moderate damping at the thickness deposited in this experiment. In contrast, the Auger lines [Fig. 2(b)] represent a more shallow information depth [8]. Thus the horizontal Pt and Si Auger signals represent a final, constant surface composition of Pt and Si. A detailed study of

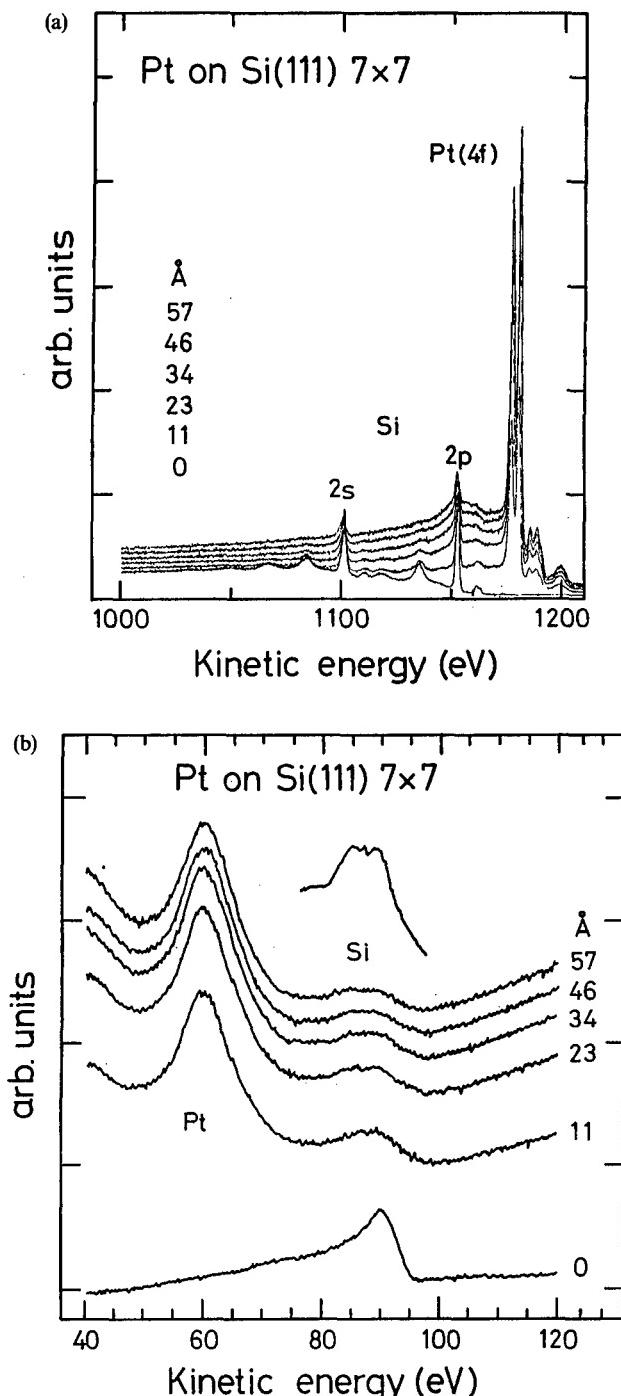


Fig. 1. (a) XPS of Pt on Si(111) 7×7 , recorded with Mg-K α radiation. (b) XAES of Pt on Si(111) 7×7 . Same conditions as Fig 1(a). The insert shows a high resolution recording of the Si(L_{2,3}VV) part of the upper spectrum, giving clear indications for silicide like features [5].

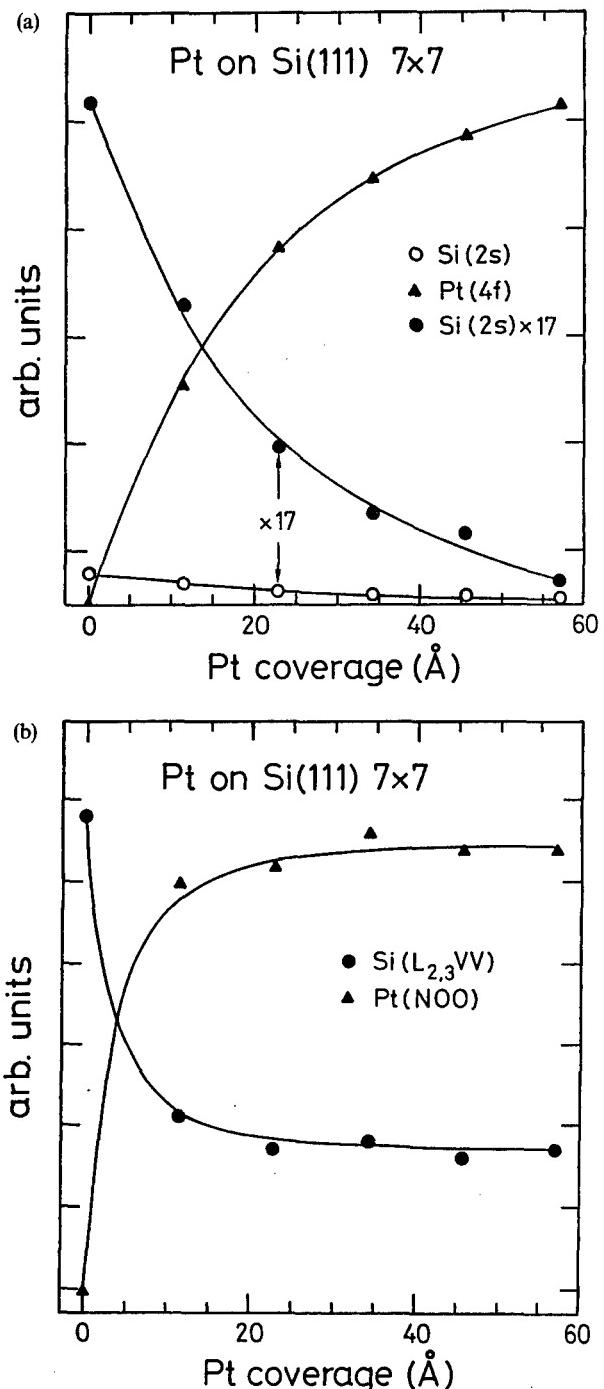


Fig. 2. (a) Si and Pt core line intensities versus Pt dose at room temperature. The filled circles represent a stretched representation of the Si(2s) signal. The triangles (Pt(4f)) and open circles refer to an identical intensity scale. (b) XAES signals [from differentiated representations of the distribution of Fig. 1(b)] versus Pt dose at room temperature. Same conditions as Fig. 2(a).

the growth of the first few monolayers is given in [3] and [4]. The small amount of islanding in this case [Fig. 2(a)] is contrasted with a larger amount observed when Pt is deposited on the surface covered by a 20 Å native oxide layer. It is detected as a deviation from a purely exponential behaviour of these plots. From the line shape of the Si(L_{2,3}VV) spectrum it is concluded that the surroundings for the surface Si atoms are similar to PtSi [4]. Thus it is concluded from these measurements that a reacted surface layer is formed with Si atoms in a Pt rich film. No signs of the buried interface are available in this case.

With the facilities at MAX Lab in Lund it is possible to do surface core level spectroscopy, by choosing an appropriate photon energy and total resolution of the system. This is demonstrated in Fig. 3, which shows the development of the Si(2p) [Fig. 3(a)] and Pt(4f) [Fig. 3(b)] core level spectra during Pt deposition. In the Si(2p) spectra four components are clearly visible to the eye without resorting to detailed

curve fitting. In agreement with the current understanding of the Si(2p) spectra of the 7 × 7 structure, the three components of the clean surface spectrum are assigned to the ad-atoms, "bulk" atoms and rest atoms of the DAS-model [9]. The displacement of the peaks with Pt coverage is indicative of chemical exchange and/or bending of the bands of the Si crystal. If the "bulk" peak is tracked, a total shift between the top spectrum and the bottom spectrum of 0.5 eV is observed. At completing the first layer the shift amounts to 0.3 eV. The shift of Pt is of the order of 0.2 eV, but in the same direction as for Si, and it occurs only after completion of the first layer. This indicates that in the reaction a net charge transfer occurs from Pt to Si. The Si band bending in the initial phase (first layer of Pt) is thus due mostly to a change of surface structure. This change is directly indicated by the changes in the Si(2p) spectrum during deposition [Fig. 3(a)], which show that the ad-atoms are removed before the first monolayer is completed. At

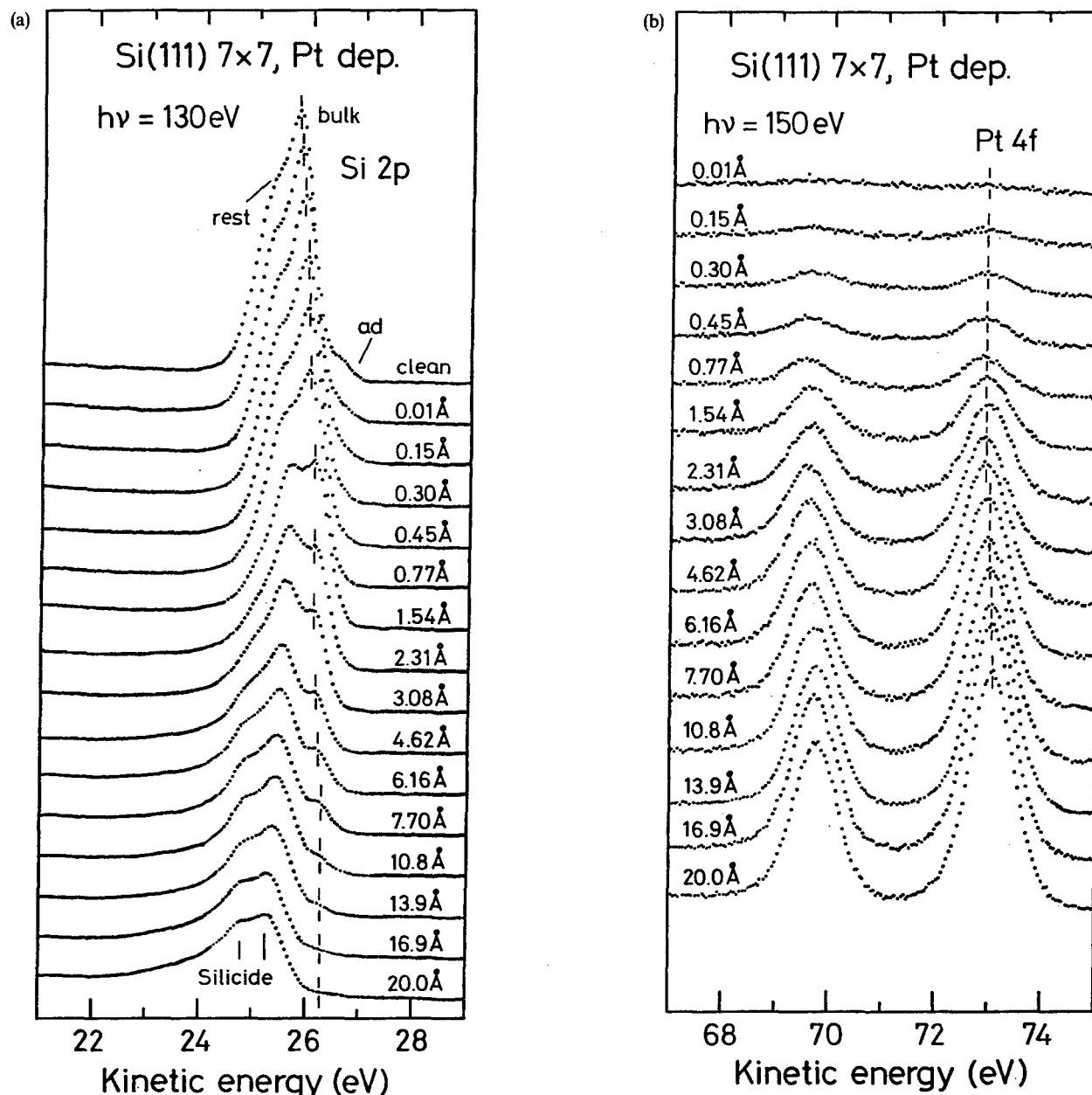


Fig. 3. (a) Si(2p) core line spectrum at $E_{\text{photon}} = 130\text{ eV}$, during Pt depositions at room temperature. See text for details. (b) Pt(4f) core line spectrum, under the same conditions as Fig. 3(a). $E_{\text{photon}} = 150\text{ eV}$.

coverages exceeding a monolayer the Si(2p) spectrum is indicative of a metallic system, characterized by the inelastic losses on the low-energy side of the peaks [10]. This is the spectrum of reacted Si and Pt (silicide). In conclusion then, a large part of the Schottky barrier height (of about 0.80 eV, corresponding to a band bending of 0.65 eV for this doping level) for the Si/Pt interface is already developed in the contact reaction between Pt and Si, without annealing. In this case, the interface is still (just) visible until the last coverage included (bulk peak), but covered with metal and silicide. The intensities of the signals are as given in [3].

Annealing studies

Results of annealing of a ~60 Å thick Pt deposit are presented in Fig. 4. The figure shows the development of the Auger spectra of Si and Pt, after various annealing cycles. It is evident that the Si ($L_{2,3}$ VV) Auger spectrum changes its shape drastically upon the first annealing cycle taking the sample up to 300 °C. This change represents a break-up of the silicide at the surface [4]. The core level intensities show that Si diffuses into the Pt film or Pt into the substrate at higher temperatures. In a separate experiment it was probed whether Pt evaporated from the sample [7]. This was found not to be the case at any of the temperatures used here. The surface concentration of Pt stabilizes after several annealing cycles up to 1200 °C. In this stage the surface is a $\sqrt{3} \times \sqrt{3}$ reconstruction with the remaining Pt atoms in sixfold interstitial sites in the top double layer of the silicon crystal [11]. Thus these experiments clearly demonstrate that a thin Pt film under these conditions will not form a homogeneous silicide with a well defined interface, but it does create a stable surface structure and a dispersed Pt concentration in the bulk of the sample.

To test some suggestions about the role of nucleation centers for the silicidation reaction, a deposition was undertaken with impurities (C and O) by not applying cooling to

the evaporator. A 15 Å thick film, heavily contaminated at the surface with C and O, was annealed, producing the XPS spectrum shown in Fig. 5. In this figure the plasmon loss peaks are depicted in more detail than before to demonstrate that a metallic like compound seems to have formed. However the surface of this sample does not represent a silicide like Si($L_{2,3}$ VV) Auger spectrum, but the ratio of the Pt and Si signals could indicate a composition like PtSi. The core level intensities of this sample also show that the Si to Pt ratio in the entire reacted phase is significantly more stable with respect to annealing than for the clean systems, at temperatures up to 950 °C. This could indicate a silicide-like bulk composition, but more work is obviously needed for this and similar cases.

Oxidation studies

The sample, which resulted from the procedures mentioned above (Fig. 5), was annealed several times at 800 °C which resulted in a clean surface with a 1×1 hexagonal LEED pattern. Room temperature oxygen exposure to 40 L (saturation of a clean Si(111) 7×7 surface [12]) does not produce any uptake of oxygen (Fig. 6).

In contrast, by keeping the sample at 500 °C during oxygen exposure, an uptake occurs as shown in Fig. 6. The oxygen signal versus exposure is shown in the insert. The saturation exposure is above 1000 L. The resulting oxide layer is thicker than for the clean surface (oxidised at room temperature) and has a Si ($L_{2,3}$ VV) spectrum as shown in Fig. 7. This spectrum (and its differentiated replica) has indications for SiO_2 like features of the oxide [13]. Also the $\sqrt{3} \times \sqrt{3}$ surface oxidises with a similar saturation exposure. The non-reacted surface does not take up oxygen at these low pressures. Thus none of the surfaces are more reactive to oxygen at room temperature than clean Si surfaces, but at elevated temperatures they are significantly more reactive than the clean surfaces.

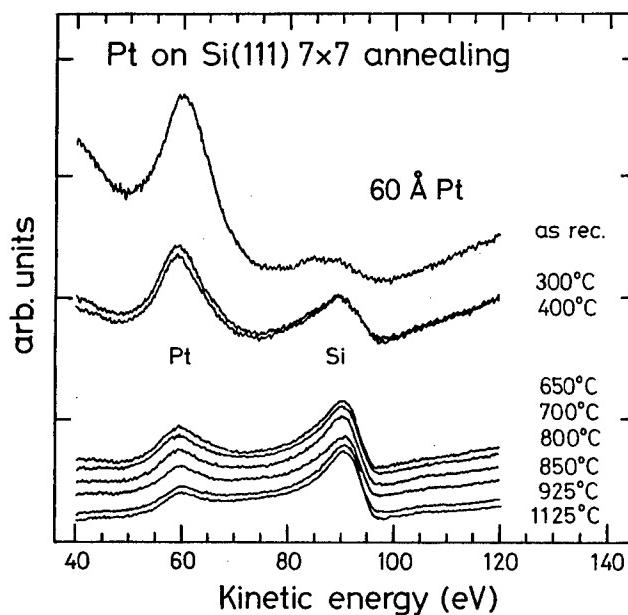


Fig. 4. XAES spectra of 60 Å Pt on Si(111) during annealing. Top spectrum: Before annealing. Spectra below represent the sequence of max. temperatures given to the right of the figure.

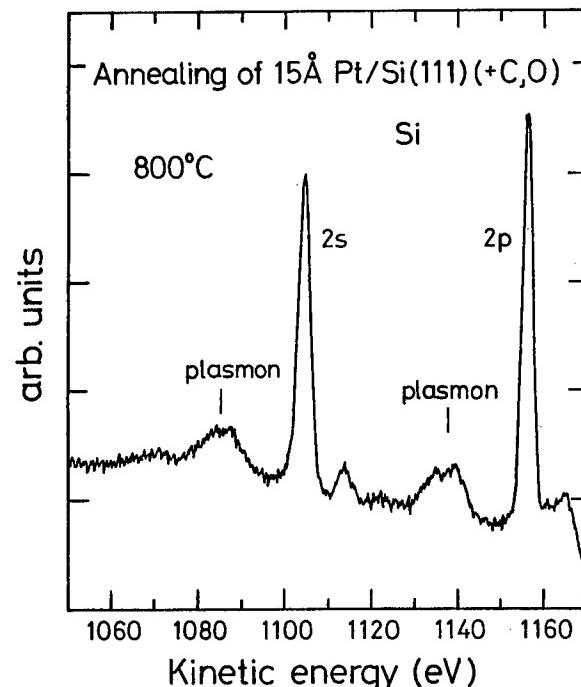


Fig. 5. XPS of Si(2s)-Si(2p) region showing plasmons after annealing of the contaminated 15 Å Pt film on Si(111) to 800 °C.

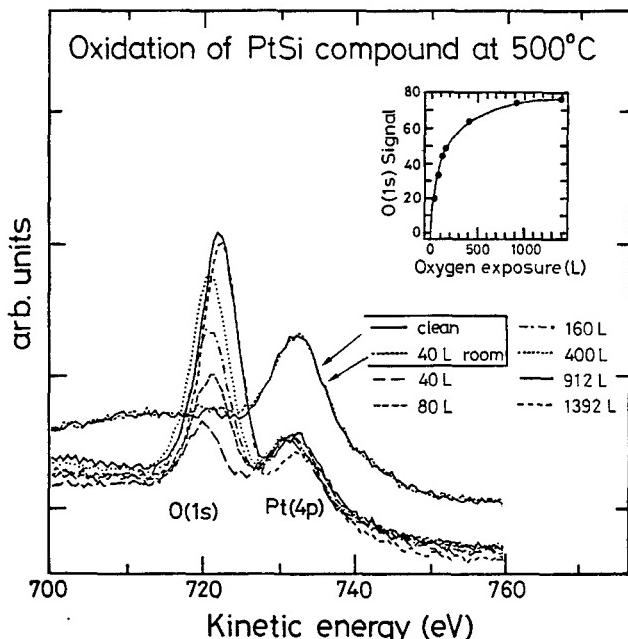


Fig. 6. O(1s) and Pt(4p) lines of a PtSi compound oxidised at 500 °C, versus oxygen exposure. The insert shows the O(1s) signal versus oxygen exposure in L (10^{-6} Torrsec).

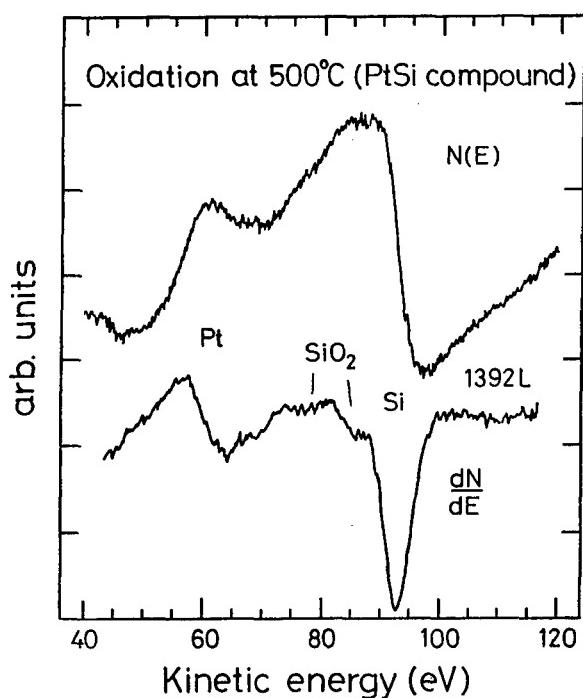


Fig. 7. XAES spectra of oxidised PtSi compound, after 1392 L oxygen exposure (Fig. 6). The lower spectrum is a derivative representation of the upper spectrum. SiO₂-like features are indicated [13].

Discussion and conclusions

This study, as a continuation of previous studies, has shown that the interaction between deposited Pt films and the Si surface occurs in several steps. The first step examined consists of the deposition of the first monolayer of Pt at room

temperature. In this step the Si ad-atoms of the 7×7 structure are gradually removed from their positions without other significant changes of the surface structure. They mix with Pt atoms after completion of a monolayer to form a local silicide-like phase at the surface. Below that phase a Pt-rich phase is growing with further Pt arriving. The slightest thermal excitation breaks up this phase and reorders the entire surface region by diffusional motion. A large part of the band bending occurs from removal of the ad-atoms in the initial phase of deposition. When silicide is formed a further band bending occurs due to the polarization of the Si-Pt bonds. Other effects on the distribution or bonding of the elements at the surface can be due to impurities. It has been demonstrated here that C and O deposited together with Pt "help" in creating a stable, homogeneous and extended phase with a composition close to a silicide, with metallic properties witnessed here by its plasmon spectrum. Oxidation at elevated temperatures of (annealed) Pt deposits is strongly enhanced compared to clean Si, and must be creating diffusional mixing in the process. These findings may be important clues to the differences between conventional technological processing which are found to form Pt-silicides at low temperatures, with the assistance of gaseous ambients, and UHV conditions, under which such processes do not occur.

Acknowledgements

The synchrotron radiation studies were done at MAX Lab, Lund, by P.M. and B.J. in collaboration with C. Jansson, J. N. Andersen and U. Karlsson. A more complete account of these studies will appear separately.

References

1. "Semiconductor Interfaces and Their Implications to VLSI Device Reliability Summary", Nishi, Y., in Proc. First Int. Symposium on Control of Semiconductor Interfaces, Karuizawa, Japan (Nov. 1993), in press.
2. Matz, R., Purtell, R. J., Yokota, Y., Rubloff G. W. and Ho, P. S., J. Vacuum Sci. Technol. A2, 253 (1984), and references therein.
3. Rossi, G., Surf. Sci. Reports, Vol. 7, Nos. 1/2, 1987.
4. Morgen, P., Szymonski, M., Onsgaard, J., Jørgensen B. and Rossi, G., Surf. Sci. 197, 347 (1988).
5. Morgen P. and Jørgensen, B., Surf. Sci. 208, 306 (1989).
6. Nylandsted Larsen, A., Chevallier J. and Pedersen, A. S., Mater. Letters 3, 242 (1985), and references therein; Chang, C-A. et al., J. Vac. Sci. Technol. A4, 841 (1986).
7. Tougaard, S. et al., to be published (private communication).
8. Seah M. P. and Dench, W., Surf. Interface Anal 1, 2 (1979).
9. "Metal Induced Surface Structures and Epitaxial Silicides on Si(111)", Christer Wigren, Thesis, Lund 1992.
10. Doniach, S. and Sunjic, M. J., Phys. C3, 285 (1970)
11. Rossi, G., Chandesris, D., Roubin, P. and Lecante, J., Phys. Rev. B34, 7455 (1986).
12. Höfer, U., Morgen, P., Wurth, W. and Umbach, E., Phys. Rev. Lett 55, 2979 (1985); Morgen, P., Höfer, U., Wurth, W. and Umbach, E., Phys. Rev. B39, 3720 (1989); Höfer, U., Morgen, P., Wurth, W. and Umbach, E., Phys. Rev. B40, 1130 (1989).
13. Ramaker, D. E. et al., Phys. Rev B19, 5375 (1979); Morgen, P. and Onsgaard, J., Surf. Sci. 99, 87 (1980).

SiC – A Semiconductor for High-Power, High-Temperature and High-Frequency Devices

E. Janzén, O. Kordina, A. Henry, W. M. Chen, N. T. Son, B. Monemar, E. Sörman, P. Bergman, C. I. Harris, R. Yakimova, M. Tuominen, A. O. Konstantinov, C. Hallin and C. Hemmingsson

Department of Physics and Measurement Technology, Linköping University, S-581 83 Linköping, Sweden

Received June 6, 1994; accepted June 15, 1994

Abstract

SiC has in comparison with Si superior basic properties for applications in high-power, high-frequency and high-temperature electronics. The potential applications of SiC were known decades ago, but the poor quality of the material produced at that time has delayed the device development. However, during the last years the crystal growth process of SiC has been improved considerably. We will present some important properties of SiC, describe the two most common growth processes and discuss fundamental materials problems that remain to be solved. A further aspect, which we will discuss, is the polytypism of SiC, which may allow us to obtain generic knowledge of, for instance, defects in semiconductors.

1. Introduction

SiC has in comparison with Si superior properties regarding high-power, high-frequency and high-temperature electronics. The material has extremely high thermal conductivity (5 W/cm K), can withstand high electric fields (4 MV/cm) before breakdown also high current densities. The high bandgap results in low leakage current even at high temperatures ($> 700^\circ\text{C}$). The above mentioned properties make SiC promising as a power device material. The electrotechnical industry, with applications at high voltage could thus in the future advantageously replace Si power transistors, thyristors and rectifiers with SiC devices.

Due to its high saturated electron velocity ($2.5 \times 10^7 \text{ cm/s}$) SiC is also capable of generating large power at high frequencies. This will prove useful for radars and SiC devices could replace electronic tubes as microwave source in many applications.

One important high-temperature SiC application is the car industry, where it sometimes is advantageous to place the electronics close to the engine. A further advantage is the high chemical stability of SiC, which enables different types of sensors to work in such a chemically aggressive high temperature environment. The mining industry has similar needs for logging measurements and control.

The potential applications of SiC were known decades ago, but the poor quality of the material has delayed the device development. However, during the last years the crystal growth process of SiC has been improved considerably. Today 30 mm wafers of reasonable quality are commercially available. The availability of such wafers has made it possible to develop the SiC Chemical Vapour Deposition (CVD) process faster. This is the background for the dramatically increased, recent interest for SiC.

In this paper we will present some important properties of SiC, describe the two most common growth processes and discuss fundamental materials problems that remain to be

solved. A further aspect, which we will discuss, is the polytypism of SiC, which may allow us to obtain generic knowledge of, for instance, defects in semiconductors. The emphasis will be on the recent work performed at Linköping University and for earlier work we refer to Refs [1–13].

2. Some important properties of SiC

2.1. The polytypes of SiC

SiC is the classic example of a one-dimensional polymorphism called polytypism. The polytypes of SiC differ from one another only in the stacking sequence of double layers of Si and C atoms. Each double layer consists of a plane of close-packed Si atoms over a plane of close-packed C atoms, where each Si atom lies directly above each C atom. The succeeding double layer is stacked in a close-packed arrangement that allows for one of three possible positions (A, B or C) relative the previous double layer, see Fig. 1. Depending on the stacking sequence, various structures (i.e. cubic (C), hexagonal (H), or rhombohedral (R)) are formed. The stacking direction is the c -axis in the hexagonal frame of reference.

There really are only three polytypes of practical importance, 3C, 6H and 4H. The number denotes the number of double layers in a stacking repeat sequence and the letter designates the structure. Thus, we have 3C for cubic SiC (β -SiC), which has the same zincblende cubic lattice as Si and GaAs and is the only cubic polytype. All of the other polytypes are known as α -SiC, the most common being 6H SiC. All the polytypes, including 3C, have nearly the same interatomic distances although their bandgaps differ about 1 eV.

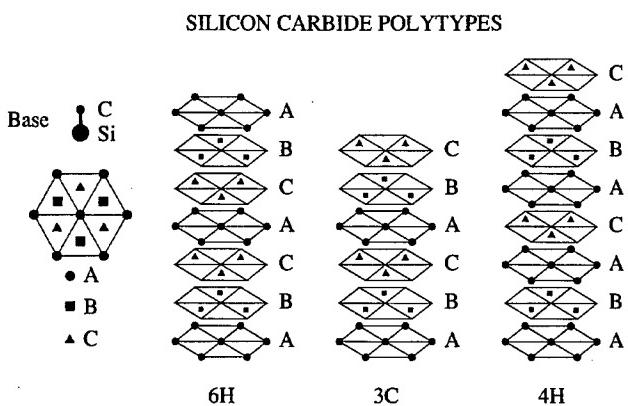


Fig. 1. The stacking sequences of the most important SiC polytypes.

2.2. Photoluminescence characterization of SiC

Substitutional donors and acceptors have more complex electronic and vibrational properties in SiC than in other semiconductors such as Si. The most common impurity in SiC is nitrogen and it is also used for n-type doping. Even though N occupies only C-sites in SiC it has one, two and three inequivalent sites to choose between in 3C, 4H and 6H-SiC, respectively, each site giving rise to a separate donor level in the bandgap. The near bandgap low temperature photoluminescence (PL) spectrum of 3C, 4H and 6H SiC is normally dominated by sharp lines due to the recombination of excitons bound at these nitrogen donors, as described below, and the energetic positions of these lines sensitively depend on the particular polytype.

Figure 2 compares two PL spectra taken from two 3C-SiC CVD epilayers; one grown on an on-axis 6H-SiC substrate (curve a) and the other on a Si substrate (curve b). E_{gx} indicates the exciton energy gap which is about 2.389 eV at 4.2 K [14]. The PL spectrum from 3C-SiC grown on 6H-SiC contains in the near bandgap region the no-phonon (NP) line of the N bound exciton (BE) at about 2.378 eV, as well as various phonon replicas [14, 15]. The binding energy of the nitrogen BE is 10 meV as measured in bulk 3C-SiC. When Si is used as the substrate [Fig. 2(b)] the NP line of the nitrogen BE from the 3C-SiC epilayer is not observed in our case and the phonon replicas are detected with a red shift and a broadening of their line widths. This behaviour is related to strain effect existing in the CVD 3C film. The strain is due to the large mismatch of about 20% in the lattice constants between Si and 3C-SiC. It was shown that for films thicker than 3 μm the film stress decreases slightly with increase of film thickness [16]. From infrared absorption the ionization energy of the N donor has been found to be 54.2 meV [17].

The hexagonal 4H polytype of SiC has an excitonic bandgap of 3.265 eV at 4.2 K [18] and Fig. 3 shows the PL spectrum in the near bandgap region of a bulk crystal. There are two NP lines marked as P_0 and Q_0 from the two inequivalent nitrogen sites, and with each NP line a series of lines is associated through phonon emission. The exciton binding energies are 7 meV and 20 meV, respectively. From infrared absorption the ionization energies of the N donors

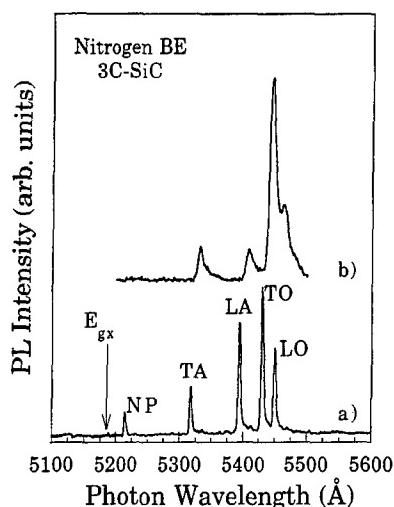


Fig. 2. PL spectra at 4.2 K showing the nitrogen BE from 3C SiC CVD films grown on (a) 6H-SiC substrate and (b) Si substrate, respectively.

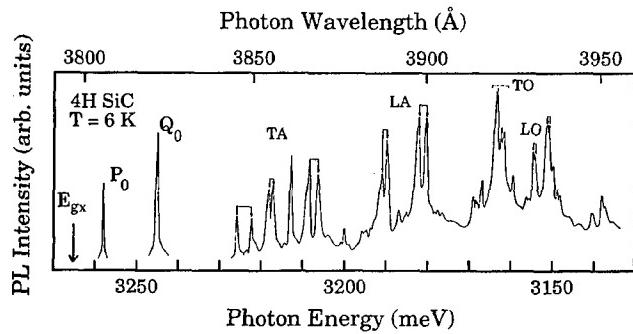


Fig. 3. PL spectrum at 6 K of the nitrogen BE complexes in 4H SiC. P_0 and Q_0 are the NP lines due to two inequivalent nitrogen sites. Other labeled lines are one-phonon replica lines of the P series. Weaker lines are Q series or intrinsic (from [18]).

are found to be 51.8 meV and 91.4 meV for the two sites [19].

Nitrogen in 6H-SiC gives rise to three donor levels at 81 meV, 137.6 meV and 142.4 meV below the conduction band, respectively [20]. Figure 4 shows PL spectra recorded at 4.2 K from different CVD epilayers grown on 6H-SiC: (a) high-doped n-type ($2 \times 10^{16} \text{ cm}^{-3}$), (b) a low-doped n-type ($5 \times 10^{14} \text{ cm}^{-3}$), (c) a low-doped p-type ($5 \times 10^{14} \text{ cm}^{-3}$) and (d) an extremely low-doped p-type ($< 1 \times 10^{14} \text{ cm}^{-3}$) sample, respectively. These spectra contain several sharp lines related to the recombination of the excitons bound to the three nitrogen donors, to the three aluminium acceptors and to the recombination of the free excitons (FEs). No-phonon BE lines (P_0 , R_0 and S_0) associated with the three nitrogen donor levels are observed together with their phonon replicas. The energy position of these three NP BE lines corresponds to the BE binding energies of 16.0 meV, 30.6 meV, and 32.4 meV, if the exciton band gap is taken as 3.023 eV [15]. For the p-type epilayer the unresolved NP lines of the Al-BE [21] are detected. Since SiC has an indirect bandgap the FE NP line is not observed. However, various phonon replicas are clearly observed and in some cases dominate the PL spectra of the low-doped CVD epilayers.

Other defect centers can be observed in the PL spectra of the different SiC polytypes. The PL spectrum of the Ti

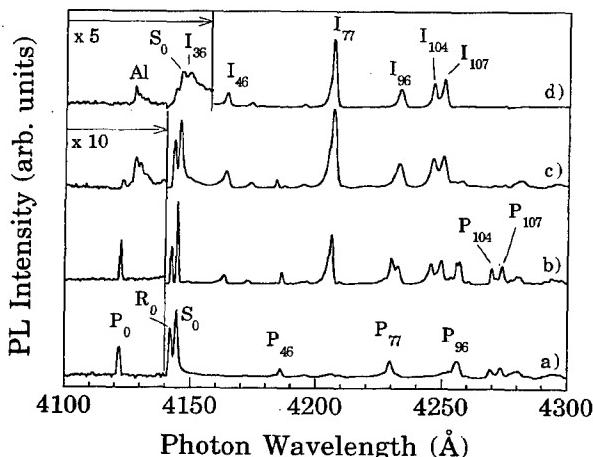


Fig. 4. PL spectra at 4.2 K of four different CVD 6H epilayers. (a) is a high-doped n-type, (b) is a low-doped n-type and (c) and (d) are low-doped p-type films. P_0 , R_0 and S_0 are the NP lines due to the three inequivalent N sites. Al denotes the NP unresolved lines due to the Al BE. I_h is the label for the intrinsic (or FE) lines, where h is the phonon replica energy.

impurity has been observed in various polytypes and the NP line energies were shown to be almost independent of the polytype energy gap. The experimental results were explained with a model in which Ti is an isoelectronic substituent for Si [22]. Radiation defects introduced by ion-implantation or electron bombardment give rise to various PL spectra. Some of the defects persist after a 1600 °C anneal and they are also observed in CVD grown material. They are mainly the D_i center in 3C and 6H CVD epilayers and also the D_{II} center in 3C film [16, 23]. PL lines due to hydrogen related centers, similar to hydrogen spectra of implanted and annealed Lely samples, have also been seen in 4H and 6H CVD films [24]. Visible luminescence due to donor (N)-acceptor (Al, Ga, B) pair recombination is often observed in heavily doped material [25]. Finally vanadium was shown to be an amphoteric deep level defect substituting Si in 4H and 6H polytypes and associated with infra-red luminescence [26].

3. SiC Growth

Here we will describe only the two most important growth techniques, the seeded sublimation technique (modified Lely) for bulk growth and the CVD technique for layer growth. For older bulk growth techniques like Acheson and original Lely we refer to [27]. Other layer growth techniques are Liquid Phase Epitaxy (LPE) [28, 29], gas source Molecular Beam Epitaxy (MBE) [30] and magnetron sputtering [31]. The latter technique has recently been proven to give high-quality 3C SiC on Si substrates [31].

3.1. Seeded sublimation growth of bulk material

Seeded sublimation growth of bulk SiC is based on material transport by sublimation from a hot source of SiC to a SiC seed crystal at lower temperature in a quasiclosed volume [32]. Thus, by employing a seed, control of the nucleation and polytype homogeneity becomes possible.

Sublimation growth set-up (Fig. 5) is typically designed as a cylindrical reaction chamber, equipped with an RF inductive heating system. In the chamber a crucible assembly is placed, consisting of a graphite lid on which a seed is set, and a cylindrical graphite crucible containing the source material. The crucible assembly is covered with heat insulators made of graphite felt. Since the growth temperature is

around 2200 °C and the source is heated up to 100–200 °C more, some parts of the growth equipment require water cooling. When doping is intended the dopants are either added to the starting material or introduced as gas mixtures.

This type of equipment is most commonly used for producing 6H and 4H SiC ingots with a growth rate of about 1 mm/h. The typical boule size is 30 mm in diameter and several cm in length.

In fact, the SiC bulk growth from vapor phase is to be regarded as a superposition of several interrelated processes such as evaporation of the starting material, heterogeneous nucleation, mass transport to the growing surface, and specific heat-exchange mechanisms. Consequently, the essential growth parameters are the growth temperature, the axial temperature gradient, the gas pressure and the source to seed distance. The growth rate of the bulk crystals is mainly governed by the temperature and pressure conditions, although experimental evidence of source to seed distance effect has also been noticed [33].

The SiC sublimation growth can take place within a wide temperature range, 1700–2500 °C. However, it has to be noticed that the growth temperature together with the seed determine the polytype and orientation of the grown material. The axial temperature gradient in the growth zone varies from 20 to 40 °C/cm depending on the growth conditions chosen. This process characteristic is directly related to the growth rate as the growth species are transferred by diffusion from the source to the seed. The mass transfer is also affected by the fluxes of the gas phase components, i.e. Si, Si₂C and SiC₂. Si is acting as C transport agent and the Si vapor pressure is the limiting one for setting the growth pressure at T < 2400 °C. The total gas pressure in the growth chamber can vary from several Torr to 1 atm of Ar-gas. The low pressure alleviates the mass transfer, i.e. the growth rate [33].

The crystal structure quality depends on the substrate perfection and on the vapor supersaturation which in its turn is determined by the temperature and the difference between the source and the seed temperature. The purity of the grown material is strongly affected by the impurities in the starting SiC which can be either purified abrasive powder or synthesized polycrystalline material [33, 34].

3.2. CVD growth of epitaxial layers

Growth of high-quality epitaxial layers at a reasonable growth rate (1–3 µm/hr) is generally done by CVD. In conventional atmospheric pressure CVD, the most common precursor gases are silane and propane. These are diluted in a massive flow of carrier gas which normally is hydrogen. The precursors are made to decompose in the hot zone of the reactor and an epitaxial film will grow on the substrate, see Fig. 6. The growth process is made at high temperatures (900 °C–1700 °C sometimes even higher) where the major part of the thermal losses are in the form of radiation. The reactor used at Linköping University has reduced these losses greatly due to a unique hot-wall susceptor design [35].

The purity of commercially available propane is no better than 99.95% (20 ppm nitrogen), which in many cases is sufficient. However, the nitrogen content in this gas is far too high to realize ultra-pure material. The purest available

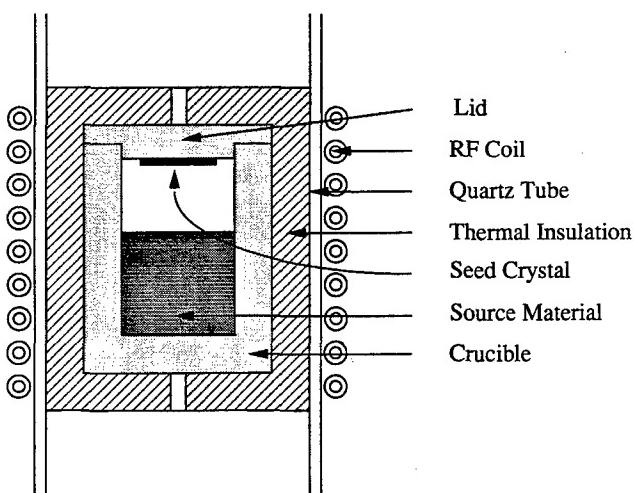


Fig. 5. A schematic cross-sectional view of SiC sublimation growth system.

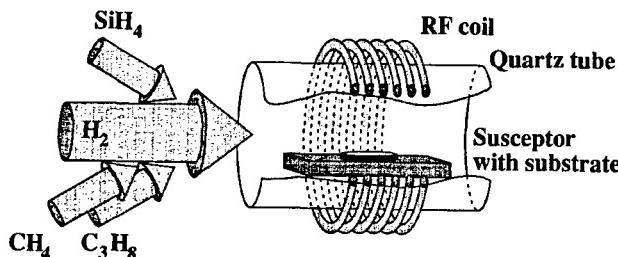


Fig. 6. A schematic view of a SiC CVD reactor.

hydro-carbon gas is methane with a nitrogen content of 0.5 ppm. This appears to be an obvious choice for growth of high purity layers but the decomposition energy for methane is much higher than that for propane. The hot-wall reactor we use allows us to grow with methane as precursor gas since the decomposition efficiency in the reactor is high. The ratio between the concentration of carbon vs silicon (C/Si ratio) needs to be much higher when methane is used. Additionally, we notice a slightly worse morphology on the layers produced with methane [36].

Growth on different substrates yields different results. Growth on Si substrates always gives the 3C polytype. Growth on Si substrates is preceded by a carbonization procedure developed by Nishino *et al.* [37]. The growth temperature when Si is used as a substrate can not be too high since Si melts just above 1400 °C. A temperature around 1200 °C has proved suitable for growth of 3C on Si substrates in our reactor [35]. Growth on on-axis 6H SiC substrates give 3C, if a high growth rate together with a low temperature is used. The 6H polytype can be produced at as low temperatures as 1200 °C with a very low growth rate

[35]. The off-axis 6H substrates normally always give the 6H polytype, since the steps guide the atoms into the right position [38].

4. Fundamental physics problems to be solved

The different polytypes of SiC can serve as a model system. They have the same kind of atoms, the same kind of bonding between the atoms and almost exactly the same interatomic distances. Only the stacking sequences differ. But this affects the band structure and the phonon dispersion relations and thus all electronic and vibrational properties.

To obtain generic knowledge about defects in semiconductors it is advantageous to compare the properties of the "same" defect both at inequivalent sites in the same polytype and at sites in different polytypes. In the next section we will report such an investigation on the shallow donor nitrogen.

4.1. Luminescence decay of nitrogen bound excitons

We have performed time-resolved measurements on the nitrogen BEs in both 3C and 6H SiC using the time correlated photon counting technique [39]. The decay time of the three different excitons in 6H SiC has been measured separately at 2.0 K and is found to be 1.5 ± 0.1 ns, 1.8 ± 0.1 ns and 8.0 ± 0.5 ns, for the S, R, and P, respectively. The corresponding decay curves are shown in Fig. 7(a). The observed decay times for the donor BE's in 6H do not differ between samples with different doping levels, in the range of 10^{16} to 10^{18} cm^{-3} .

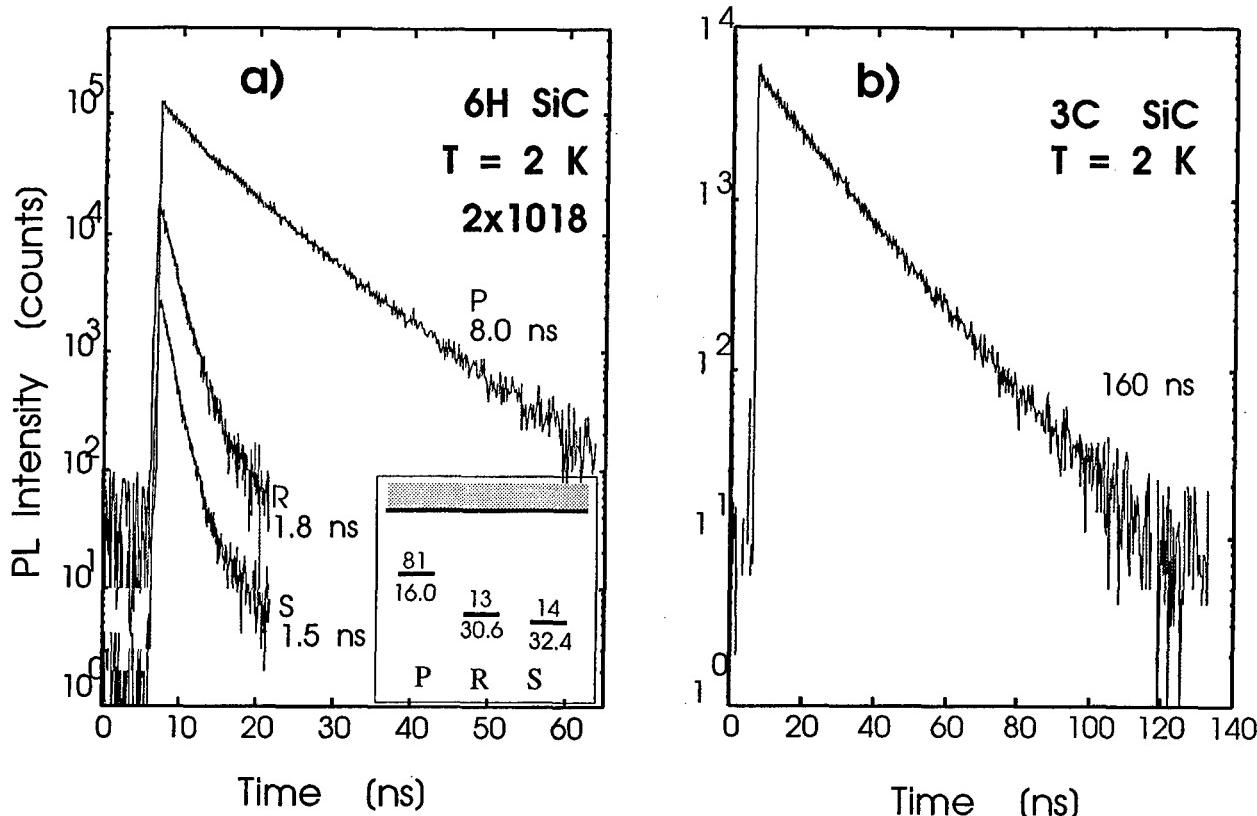


Fig. 7. (a) Photoluminescence decay curves at 2 K of the three nitrogen BEs in 6H SiC. The inset shows the corresponding donor energy level (top) and the binding energy of the exciton (bottom). The decay times, which are extremely fast, are shown for each exciton. (b) Photoluminescence decay curves at 2 K of the nitrogen BE in 3C SiC. The decay time is found to be 160 μs.

We have also measured the temperature dependence of the decay time in 6H SiC. The measured decay time is found to be constant up to a certain temperature, above which it rapidly decreases. This can most simply be described with the introduction of an additional temperature activated non-radiative recombination channel. The activation energy for this recombination is within experimental errors equal to the binding energy of the BE. We therefore conclude that the decrease of the observed decay time is due to the thermal ionization of the BE into a FE.

The observed decay time at 2 K of the N BE in 3C SiC is considerably longer, 160 ± 5 ns, as shown in Fig. 7(b). Also in this case, we have obtained the same value for the decay time in several samples, in both bulk and epitaxial layers.

The observed decay times in both 6H and 3C are considerably faster than expected for a radiative recombination in a semiconductor with an indirect bandgap like SiC. It appears that non-radiative recombination is the dominant mechanism that reduces the BE population. From the results of these decay time measurements – i.e. fast recombination, which is temperature independent at low temperatures, no dependence on doping concentration, correlation between decay time and energy level of the corresponding donor – we conclude that the most probable non-radiative process is a phonon-less Auger process [40, 41].

The donor BE consists originally of three particles, two electrons and one hole. In the Auger process one of the electron recombines with the hole leaving the excess energy and momentum to the second electron. This electron, labelled as the Auger particle, is then excited up into the conduction band with an energy close to the bandgap energy and with an additional wave vector determined by the separation of the electron and hole in momentum space, see Fig. 8.

In 6H SiC, the hole is located at the valence band top, which is located at the Γ -point in the zone center, whereas the two electrons are located at the conduction band minimum (assumed to be at or close to the M-point at the zone edge). This means that the Auger electron receives an additional momentum corresponding to the indirect bandgap, and that the final state of the Auger electron in

k-space will be in the vicinity of the zone center. Existing bandstructure calculations [42] show that the bandgap at the Γ -point is close to or larger than twice the minimum indirect bandgap. The exact value of the zone center bandgap is in this case very important, since the recombination rate is strongly dependent on whether the final energy of the Auger particle is located below or above the Γ -point band gap. According to our measurements the recombination process is very fast (in the low ns range) and we thus conclude that the direct bandgap is less than twice the indirect bandgap.

The nitrogen donor levels in 6H SiC are relatively deep causing a strong localisation for the primarily bound electron. This will consequently lead to a delocalized wavefunction in k-space and a higher probability of Auger recombination for final state wavevectors different from $\mathbf{k} = 0$. This is further demonstrated by the observed decrease in decay time with increasing energy for the three nitrogen donor levels.

5. Fundamental material problems to be solved

There are several fundamental material problems that need to be solved before SiC devices can be of any substantial commercial value. The most severe problem is the existence of small holes, so-called micropipes, penetrating the wafers. The diameter of the micropipes may vary from below μm to tens of μm . The micropipe density is today $50\text{--}200\text{ cm}^{-2}$ in high quality material. This makes the yield for devices with areas larger than a few mm^2 extremely small. In power applications the device areas are often as large as several cm^2 . Thus, to make large area devices at all possible and to improve the yield of small area devices the micropipe density must be reduced considerably.

In order to realize high voltage rectifying properties it is necessary to produce thick low-doped active layers. For instance, a 5 kV rectifier needs a $40\mu\text{m}$ thick active layer with a controlled doping concentration in the low 10^{15} cm^{-3} range. The residual doping should be in the low 10^{14} cm^{-3} range.

For bipolar, high-voltage devices the carrier lifetime in the active layer must be long under high-injection conditions. Otherwise the forward voltage drop becomes too high. The previously mentioned 5 kV rectifier would require a high-injection lifetime of about $1\mu\text{s}$.

For high-frequency devices the frequency limit is set by the coupling to the conducting substrate. Semi-insulating substrates or thick semi-insulating films to isolate the device structures would thus be an important improvement.

We will in the following two sections report on the latest results achieved at Linköping University concerning low residual carrier concentration and long carrier lifetime.

5.1. Low residual carrier concentration

Nitrogen is, as previously mentioned, the most common defect in SiC. CVD-layers with very low nitrogen concentration can be grown, though great care must be taken prior to and during growth. Due to the high nitrogen content of the propane precursor, it is difficult to grow layers with a nitrogen concentration lower than 10^{15} cm^{-3} . Compensating acceptors can, however, unintentionally be introduced from

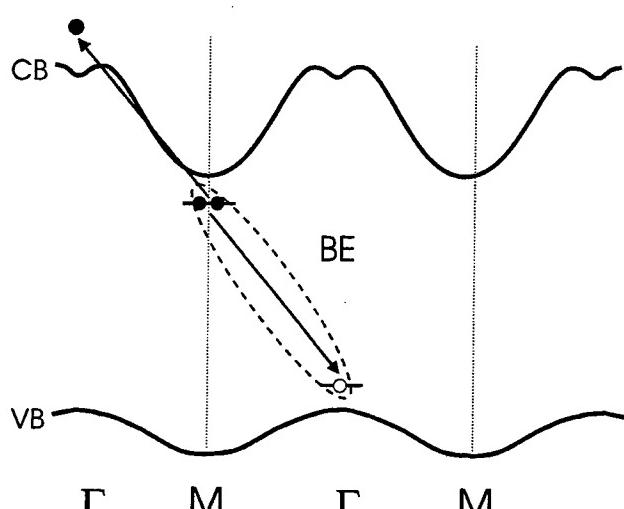


Fig. 8. A schematic view of the donor BE Auger recombination in 6H SiC. One of the electrons of the BE recombines with the hole. The excess energy and momentum is transferred to the second electron, which is excited up in the conduction band.

the graphite if the protecting SiC film on the susceptor is partly etched away by the hydrogen flow. This can be seen in PL, since Al BEs will appear in the PL spectra, see Fig. 4(c) and (d). Figure 4(b) shows a PL spectrum from a low doped n-type sample without any noticeable compensating acceptors. By comparing the ratio between the N BE and the FE (I_{77}) with the doping concentration of the layer for n-type samples, a clear correlation is seen Fig. 9. Layers with little or no compensating acceptors (i.e. no Al BEs could be observed in the PL spectra) produce a straight line in Fig. 9. Compensated films all lie above this line, since they have lower net doping concentration and due to a relatively higher amount of impurities also have lower FE intensity.

Methane is, as mentioned previously, a purer source than propane. Using methane uncompensated n-type layers have been produced with a residual doping concentration as low as $5 \cdot 10^{14} \text{ cm}^{-3}$. If the C/Si ratio is increased the nitrogen incorporation is decreased and, provided there are any compensating acceptors in the reactor, an extremely low doped p-type layer can be grown. Figure 4(d) shows a PL spectrum of such a low-doped p-type film. The Al BEs are clearly seen, however the spectrum is almost entirely dominated by FE related lines. Note that the nitrogen BE lines are merged with the I_{36} FE related line. The doping concentration is not measured on this sample but is well below 10^{14} cm^{-3} . Although methane gives much purer layers, propane is normally preferred due to the superior morphology obtained with this source. Propane was therefore the obvious choice when a structure was grown intended to be processed into a high voltage rectifier potentially capable of handling 5 kV. On a n^+ substrate, a $45 \mu\text{m}$ n-type layer (approximately $1 \cdot 10^{15} \text{ cm}^{-3}$) was grown, followed by a heavily p-type doped layer $1.5 \mu\text{m}$ thick. The p-type doping was achieved by a small addition of trimethylaluminum. The layer exhibits excellent morphology with the exception for a very small corrugation just noticeable at 400 times magnification in a Normarski-type optical microscope. This we believe is due to step bunching effects. PL shows that the active

n-layer has excellent crystalline quality with no signs of any compensating acceptors.

5.2 Long carrier lifetime and lifetime limiting defect

We have managed to grow CVD layers with reasonable long carrier lifetimes. For uncompensated layers with doping less than 10^{15} cm^{-3} the minority-carrier lifetime approaches $0.5 \mu\text{s}$ at room temperature.

Up to now very little is known about deep level defects in SiC, which play important roles in carrier recombination. Among them, the dominant recombination center (i.e. the lifetime-limiting defect), despite of its significant importance, is still a mystery. This is mainly due to the lack of suitable experimental techniques which not only are sensitive to carrier recombination but also provide detailed information about the defect electronic and geometrical structure.

Optical detection of magnetic resonance (ODMR) has in the past decade proven to be a very powerful technique, which combines highly sensitive optical spectroscopy with microscopically informative magnetic-resonance techniques [43]. ODMR is a suitable technique in studying the lifetime limiting defect, since the radiative carrier recombination processes are directly monitored in the experiments. A magnetic-resonance enhanced carrier recombination via the dominant recombination center will result in a corresponding decrease in carrier recombination via other channels [44]. In other words, the PL emission directly related to the dominant recombination channel, if it is radiative, will increase upon the magnetic resonance condition. PL emissions other than the dominant recombination center will consequently decrease due to the competing process.

The samples studied include 4H and 6H bulk crystals grown by the modified Lely method and epilayers grown by CVD. They are either n- or p-type doped. The carrier recombination processes are monitored here by PL emissions from the materials. In all these samples studied, an ODMR signal at around $g = 2$ can be observed via all the PL emissions from the samples, see Fig. 10. The resonance line at X-band frequencies is rather broad and probably contains unresolved hyperfine structure and unresolved resonances from different inequivalent lattice sites of the defect. The spectral dependence study shows that this ODMR signal corresponds to an enhancement in the intensity of the

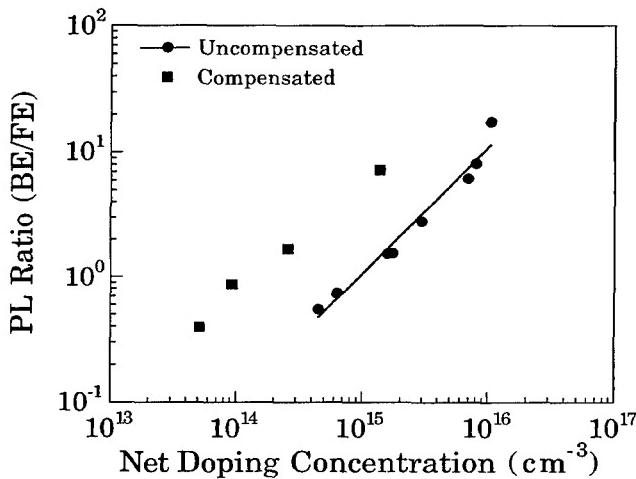


Fig. 9. The dependence of the BE/FE ratio on the net doping concentration for n-type epilayers at 4.2 K. The BE/FE ratio is determined from the PL spectra by the following equation $\text{BE}/\text{FE} = (I_R W_R + I_S W_S) I_{FE} W_{FE}$ where I_R , W_R , I_S , W_S , I_{FE} and W_{FE} are the intensities and full widths at half maximum of the R_0 , S_0 , and I_{77} lines, respectively. The net doping concentration is determined from C-V measurements.

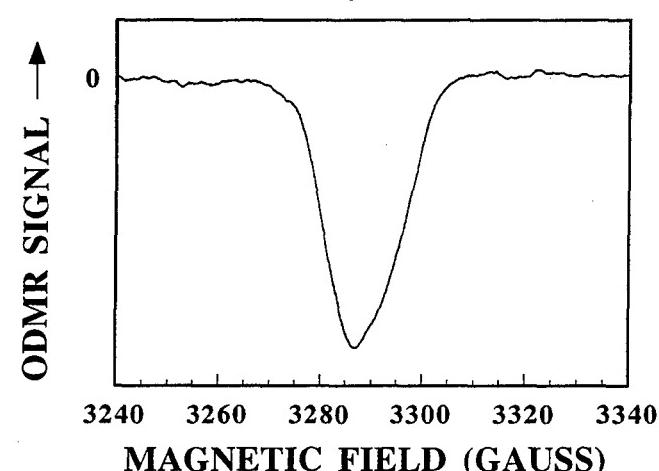


Fig. 10. ODMR spectrum obtained by detecting the entire visible PL emission of 6H SiC, at 4 K and 9.23 GHz.

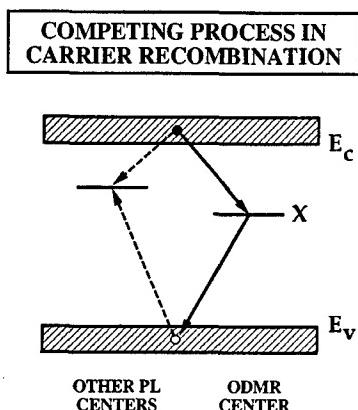


Fig. 11. A schematic picture of competing carrier recombination processes. An enhancement in carrier recombination via the lifetime limiting defect (denoted as X in Fig. 11) leads to a corresponding decrease of carrier recombination via other channels in the material.

PL band peaking at 1.65 eV and a decrease in the intensity of other PL emissions ranging from bandedge shallow BEs to near-infrared deep PL bands. This is interpreted as due to the competing process in carrier recombination as schematically shown in Fig. 11: the magnetic resonance at the dominant recombination center (denoted as X in Fig. 11) promotes carrier recombination via this center (the PL band at 1.65 eV) and consequently reduces carrier recombination via other recombination channels (represented by the dashed arrows).

This dominant recombination center is shown to be a deep level defect, evident from the related deep PL emission and a photo-excitation spectrum of the ODMR signal. The latter reveals the photo-ionization of the center, which determines the energy level of the defect to be at about 1.1 eV below the bottom of the conduction band. The fact that this ODMR signal has been observed in both 6H and 4H SiC, regardless of their doping type and whether they are bulk or epilayers, indicates that this must be a common and basic defect in SiC, probably involving native defects. This defect provides a very efficient recombination channel, limiting the carrier lifetime.

6. Summary

The progress of SiC technology has been substantial during the last years and may lead to a break-through in high-power, high-frequency and high-temperature electronics in the near future. The most important obstacle left is the micropipe density, which has to be reduced considerably ($< 1 \text{ cm}^{-2}$) before SiC electronics can be of any real commercial value. From a fundamental point of view the progress of the SiC material quality, will enable us to obtain generic knowledge of solid state physics by comparing the same property for several different polytypes.

Acknowledgements

Financial support for this work was provided by the Swedish Board for Industrial and Technical Development (NUTEK)/Asea Brown Boveri (ABB) Power Device Program, the Swedish Council for Engineering Sciences (TFR), NUTEK, the Swedish Natural Science Research Council (NFR), the NUTEK/NFR Material Consortium on Thin Film Growth and the Swedish Institute.

References

1. Silicon Carbide, a High Temperature Semiconductor, Proc. Conf. on Silicon Carbide, Boston Massachusetts, April 2–3, 1959 (Edited by J. R. Connor and J. Smiltens) (Pergamon Press, New York 1960).
2. Silicon Carbide – 1968, Proc. Int. Conf. on Silicon Carbide, University Park, Pennsylvania, October 20–23, 1968 (Edited by H. K. Henisch and R. Roy), special issue of the Materials Research Bulletin (Pergamon Press, New York 1969).
3. Silicon Carbide – 1973, Proc. 3rd Int. Conf. on Silicon Carbide, Miami Beach, Florida, 17–20 Sept. 1973 (Edited by R. C. Marshall, J. W. Faust, Jr. and C. E. Ryan) (University of South Carolina Press, Columbia, SC, 1974).
4. Amorphous and Crystalline Silicon Carbide and Related Materials, Proc. 1st Int. Conf., Washington DC, December 10 and 11, 1987, Springer proceedings in Physics vol 34 (Edited by G. L. Harris and C. Y. Yang) (Springer-Verlag, Berlin 1989).
5. Amorphous and Crystalline Silicon Carbide II, Recent Developments, Proc. 2nd Int. Conf., Santa Clara, CA, December 15–16, 1988, Springer proceedings in Physics vol 43 (Edited by M. M. Rahman, C. Y. Yang and G. L. Harris) (Springer-Verlag, Berlin 1989).
6. Amorphous and Crystalline Silicon Carbide III and Other Group IV-IV Materials, Proc. 3rd. Int. Conf., Howard University, Washington, D.C., April 11–13, 1990, Springer proceedings in Physics vol 56 (Edited by G. L. Harris, M. G. Spencer and C. Y.-W. Yang) (Springer-Verlag, Berlin 1992).
7. Amorphous and Crystalline Silicon Carbide IV, Proc. 4th Int. Conf., Santa Clara, CA, October 9–11, 1991, Springer proceedings in Physics vol 71 (Edited by C. Y. Yang, M. M. Rahman and G. L. Harris) (Springer-Verlag, Berlin 1992).
8. Silicon Carbide and Related Materials, Proc. 5th Int. Conf., Washington, D.C., November 1–3, 1993, Institute of Physics Conference Series Number XXX, Institute of Physics, Bristol, 1994 (in press).
9. Wide-band-gap Semiconductors, Proc. 7th Trieste ICTP-IUPAP Semiconductor Symposium, Trieste, Italy, 8–12 June 1993 (Edited by C. G. Van de Walle) (North-Holland, Amsterdam 1993), (Physica B volume 185).
10. Landolt-Börnstein, New Series volume III/17a (Edited by O. Madelung) (Springer-Verlag, Berlin 1982), p 132.
11. Landolt-Börnstein, New Series volume III/22b (Edited by M. Schulz) (Springer-Verlag, Berlin 1989), p 490.
12. Gmelin Handbook of Inorganic Chemistry, 8th edition, Silicon Supplement B2 (Edited by G. Kirschstein and D. Koschel) (Springer-Verlag, Berlin 1984).
13. Gmelin Handbook of Inorganic Chemistry, 8th edition, Silicon Supplement B3 (Edited by H. Katscher, R. Sangster and F. Schröder) (Springer-Verlag, Berlin 1986).
14. Nedzvetskii, D. S., Novikov, B. V., Prokofeva, N. K. and Reifman, M. B., Sov. Phys. Semiconductors 2, 914 (1969).
15. Choyke, W. J., Hamilton, D. R. and Patrick, L., Phys. Rev. 133, A1163 (1964).
16. Choyke, W. J., Feng, Z. C. and Powell, J. A., J. Appl. Phys. 64, 3163 (1988).
17. Moore, W. J. et al., Phys. Rev. B48, 12289 (1993).
18. Patrick, L., Choyke, W. J. and Hamilton, D. R., Phys. Rev. 137, A1515 (1965).
19. Götz, W. et al., Materials Science Forum 117–118, 495 (1993).
20. Suttorp, W., Pensl, G., Choyke, W. J., Stein, R. and Leibenzeder, S., J. Appl. Phys. 72, 3708 (1992).
21. Clemen, L. L. et al., Appl. Phys. Lett. 62, 2953 (1993).
22. Patrick, L. and Choyke, W. J., Phys. Rev. B10, 5091 (1974).
23. Henry, A., Kordina, O., Bergman, J. P., Harris, C. I. and Janzen, E., unpublished.
24. Clemen L. L. et al., in Ref. [8].
25. Suzuki, A., Matsunami, H. and Tanaka, T., J. Electrochem. Soc. 124, 241 (1977).
26. Dörnen, A. et al., Proc. of the 16th Int. Conf. on Defects in Semiconductors 16, Pennsylvania, Mat. Science Forum 83–87, 1213 (1992).
27. Knippenberg, W. F., Philips Res. Repts. 18, 161 (1963).
28. Yakimova, R. and Kalnin, A., Phys. Stat. Solidi (a) 32, 297 (1975).
29. Hoffman, L., Ziegler, G., Theis, D. and Weyrich, C., J. Appl. Phys. 53, 6962 (1982).
30. Rowland, L. B., Kern, R. S., Tanaka, S. and Davis, R. F., J. Mater. Res. 8, 2753 (1993).

31. Wahab, Q. *et al.*, unpublished.
32. Tairov, Y. M. and Tsvetkov, V. F., *J. Crystal Growth* **43**, 209 (1978).
33. Barett, D. L. *et al.*, *J. Crystal Growth* **128**, 358 (1993).
34. Hobgood, H. M., McHugh, J. P. and Hopkins, R. H., in Ref. [8].
35. Kordina, O., Hallin, C., Glass, R. C., Henry, A. and Janzén, E., in Ref. [8].
36. Kordina, O. *et al.*, in Proc. MRS spring meeting 4–8 April, 1994, San Francisco, in press.
37. Nishino, S., Suhara, H., Ono, H. and Matsunami, H., *J. Appl. Phys.* **61**, 4889 (1987).
38. Kimoto, T., Nishino, H., Yoo, W. S. and Matsunami, H., *J. Appl. Phys.* **73**, 726 (1993).
39. O'Connor, D. V. and Phillips, D., "Time-correlated photon counting" (Academic Press, London 1984).
40. Schmid, W., *Phys. Stat. Sol.* **84**, 529 (1977).
41. Osbourn, G. C. and Smith, D. L., *Phys. Rev. B* **16**, 5426 (1977).
42. Gavrilenko, V. I., Postnikov, A. V., Klyui, N. I. and Litovchenko, V. G., *Phys. Stat. Sol.* **162**, 477 (1990).
43. Cavenett, B. C., *Adv. in Physics* **30**, 475 (1981).
44. Chen, W. M. and Monemar, B., *Appl. Phys. A* **53**, 130 (1991).

Thermal Oxidation of n- and p-type 6H-Silicon Carbide

C.-M. Zetterling and M. Östling*

Royal Institute of Technology, Department of Electronics, Solid State Electronics, Electrum 229, S-164 40 Kista-Stockholm, Sweden

Received June 2, 1994; accepted June 15, 1994

Abstract

Thermal oxides have been grown on monocrystalline 6H silicon carbide samples (n-type and p-type) with both carbon face and silicon face. The oxidation was performed in a dry oxygen ambient at 1523 K with or without the addition of TCA (Trichloroethane), or in wet pyrogenic steam at 1473 K. Polysilicon gates doped with POCl_3 were used for electrical characterisation by capacitance-voltage measurements at room temperature. Large flatband voltage shifts indicate fixed oxide charges up to 10^{13} cm^{-2} .

1. Introduction

Silicon carbide, SiC , has been suggested as a suitable material for high temperature and high power electronic devices [1, 2, 3], due to its wide energy bandgap and high thermal conductivity. SiC exists in several crystal structures, and the hexagonal structures may appear in several polytypes. Three of these are commercially available: 6H (hexagonal), 3C (cubic), (grown on (100) silicon) and recently 4H (hexagonal). When polishing the wafers, either the (0001) silicon face or the (0001̄) carbon face may be selected, see Fig. 1 [4]. Many properties of the material differ depending on which face is used. The oxidation rate is higher on the carbon face, see Table I and [5, 6], and also CVD homoepitaxy is faster on on-axis carbon face wafers [7].

Oxidation studies have received much attention, since high quality oxides will be needed for MOS-gates, insulation and passivation. Most oxidation studies so far on 6H SiC have concentrated on the kinetics, and on n-type material (nitrogen doped) [6, 8]. Although the oxide quality has been found satisfactory on n-type, further investigations are needed on p-type substrates (aluminium doped), since difficulties have been reported in obtaining high quality oxides [9]. It is suspected that the large amounts of aluminium

detected in the oxides after thermal oxidation [10] (and [11] on 3C SiC) will influence both the threshold voltage and the gate oxide breakdown of MOSFETs. In this study we compare electrical characteristics of oxides grown on the silicon face (Si face) and the carbon face (C face) with either n-type (nitrogen) or p-type (aluminum) doping.

2. Experimental

The starting material was monocrystalline 6H silicon carbide samples with the silicon face or the carbon face polished (from Cree Research Inc.). Wafers were supplied with at least $3 \mu\text{m}$ thick epilayers to reduce the concentration of defects and accurately control the doping. The initial concentrations of aluminum and nitrogen are listed in Table I. The uniform net doping levels were measured by capacitance-voltage measurements using an Hg probe.

Three different procedures were used for the oxidation:

1. Dry oxidation at 1523 K
2. Dry oxidation at 1523 K with TCA (Trichloroethane) to yield 2–4% chlorine
3. Wet Oxidation in pyrogenic steam at 1473 K

The Si face wafers were oxidized for 50 minutes and the C face for 10 minutes, in order to obtain comparable oxide

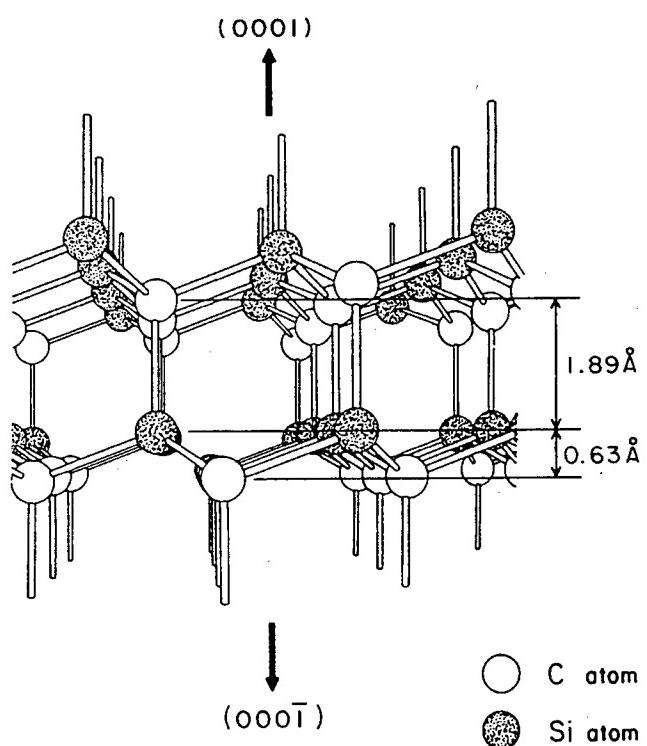


Table I. Sample data and oxidation results

CREE specifications:				
Wafer type:	n-type silicon	p-type silicon	n-type carbon	p-type carbon
Polished face:	silicon	silicon	carbon	carbon
Net doping (cm^{-3})	$8.7 \cdot 10^{15}$	$1.3 \cdot 10^{15}$	$5.8 \cdot 10^{15}$	$5.5 \cdot 10^{15}$
Oxidation time (min):	50	50	10	10
Oxide thickness (nm):				
Dry (1523 K)	44	40	51	48
Dry + TCA (1523 K)	43	42	49	50
Wet (1473 K)	40	48	130	131

* Present address: Stanford University, CIS, Stanford, CA 94305, U.S.S.
E-mail: carl-mikael@ele.kth.se (Carl-Mikael Zetterling); ostling@ele.kth.se (Mikael Östling)

Fig. 1. The silicon carbide crystal structure, and the definition of the (0001) silicon face and the (0001̄) carbon face. (Figure taken from Ref. [4])

thicknesses. After oxidation an anneal was performed in nitrogen for 60 minutes at the same temperature as the oxidation, in the same furnace. The n-type and p-type samples were oxidized simultaneously, and silicon (100) samples were used as control wafers to monitor the oxidation process. The oxide thickness was measured using ellipsometry, with a wavelength of 632.8 nm and an incident angle of 70° (the index of refraction for SiC was assumed to be 2.636), see results in Table I.

The gate electrodes were formed by depositing 450 nm polysilicon in a LPCVD process, and doped using POCl_3 , followed by e-gun evaporated aluminum to a thickness of 200 nm. Standard photolithography was used to pattern the gates. Wet etching in phosphoric acid was used to etch the aluminum, and the polysilicon was dry etched in CF_4 at 15 mTorr pressure. Circular contacts with diameters of 100 μm , 200 μm , 400 μm and 1000 μm were thus formed. The sample backsides were coated with aluminum (p-type) and nickel (n-type). Prior to electrical characterisation annealing was performed at 774 K in a vacuum furnace and at 673 K in forming gas (N_2 with 10% H_2).

3. Results

3.1. Oxide thickness

As can be seen in Table I, all oxide thicknesses were between 40 and 50 nm, except the wet oxidized C face samples which were 130 nm thick. This shows that oxide growth at these temperatures is diffusion limited on the C face and reaction limited on the Si face. A possible explanation for the dependence on crystal face may be found when comparing this to results from oxidation of silicon with different orientations [12, 13]. Roughly, fewer bonds per area at a given time leads to a higher activation energy, although steric hindrances have to be taken into account as well. Considering oxidation of the C face, see Fig. 1, it is seen that the silicon atoms will display three available bonds on the face that is being oxidized, whereas on the Si face the silicon atoms display only one bond. This is in qualitative agreement with measurements of the activation energy for oxidation, 26 and 80 kcal/mol for C face and Si face respectively [14, 15], and explains why oxide grows faster on the C face.

3.2 Electrical characterisation

Capacitance-voltage measurements were made at room temperature and in darkness using a HP4280A CV-meter with 1 MHz measuring frequency. Voltage sweeps were made from inversion/depletion to accumulation, in equilibrium, and from accumulation back to inversion/depletion. Some hysteresis was seen when the gate voltage was swept in both directions, see Fig. 2, which may be explained by charge injected into slow traps during the equilibrium sweep. To ensure that the series resistance was of minor importance, the oxide capacitance of 100 μm , 200 μm , 400 μm and 1000 μm diameter capacitances was measured and found to scale with the gate area. CV-curves for the different oxides with 400 μm diameter polysilicon gates can be seen in Figs 2–5, with capacitance values normalised to that of the oxide (200 μm gates were measured as well, and the curves were similar but noisier). Theoretical curves were calculated using the method by Brews [16], with oxide thickness, gate area and substrate doping as input parameters. The work-

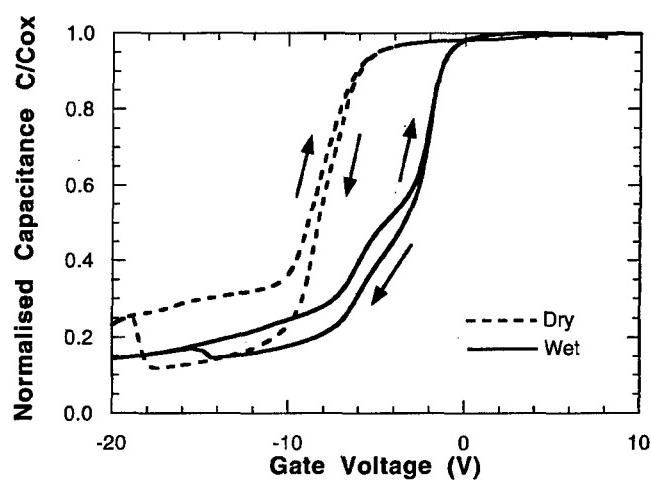


Fig. 2. Normalised capacitance vs. gate voltage curves for n-type silicon face SiC MOS capacitors. The gate area is $1.3 \cdot 10^{-3} \text{ cm}^2$. The hysteresis displayed when the gate voltage was swept in both directions is attributed to injected charge. Arrows indicate the voltage sweep direction.

function difference between gate and semiconductor, however, was set to 0 eV. Separate theoretical curves were obtained for the dry (50 nm) and wet oxides (130 nm) on the C face (Fig. 4).

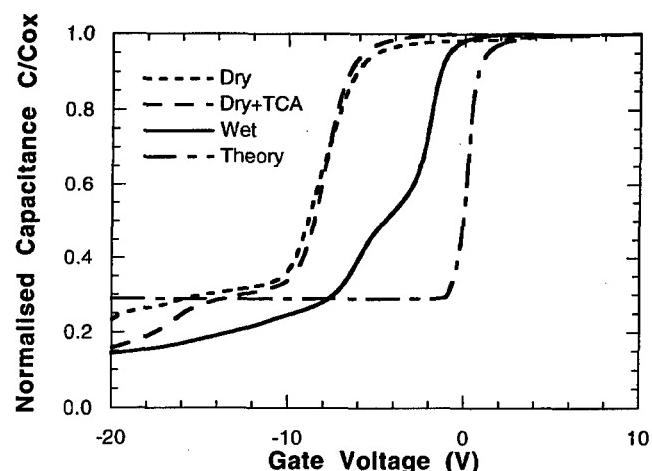


Fig. 3. Normalised capacitance vs. gate voltage curves for n-type silicon face SiC MOS capacitors. The gate area is $1.3 \cdot 10^{-3} \text{ cm}^2$.

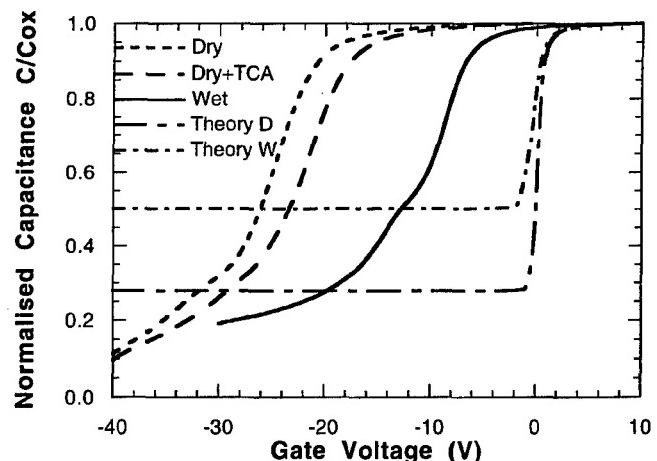


Fig. 4. Normalised capacitance vs. gate voltage curves for n-type carbon face SiC MOS capacitors. The gate area is $1.3 \cdot 10^{-3} \text{ cm}^2$.

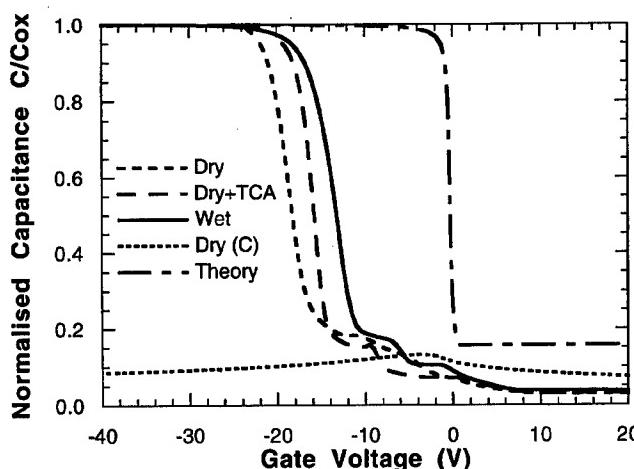


Fig. 5. Normalised capacitance *vs.* gate voltage curves for p-type silicon face and carbon face SiC MOS capacitors. The gate area is $1.3 \cdot 10^{-3} \text{ cm}^2$.

The oxide thickness measured by ellipsometry agreed well with the oxide thickness which was calculated from the oxide capacitance measured in accumulation, except for the p-type C face oxides. In this case the series resistance was too high for accurate measurement of the oxide capacitance, see Fig. 5. This problem has recently been reported by Ouisse *et al.* [9], with the remedy of measuring at low frequencies. The low capacitance part of all the curves do not display complete inversion if compared to the theoretical curves. Instead deep depletion occurs, which may be attributed to the wide energy bandgap of silicon carbide, with the resulting low thermal generation of minority carriers. This has been reported previously in literature [8, 15, 17].

From the large flatband voltage shifts in the curves in Figs 2–5 the fixed oxide charge can be estimated to be between 10^{12} and 10^{13} cm^{-2} which is extremely high. From Figs 3 and 4 it can be seen that the wet oxides display a smaller flatband voltage shift than the dry oxides on n-type substrates, but this effect is less obvious on the p-type, Fig. 5. The region with a large slope (onset of depletion) is a good indication of surface states when different curves with the same oxide thickness are compared. The larger the slope is, the less interface states are present. In Figs 3–5 the dry oxidation with added TCA displays less surface states than a standard dry oxide. The wet oxide on n-type Si face (Fig. 3) also exhibits some possible surface states.

The oxides were unfortunately very leaky, especially the p-type oxides, which made CV-measurements difficult to perform. Consequently, proper breakdown field measurements of the oxides could not be undertaken. Even at 100 Volts some oxides would not break, since the leakage current was around $1 \mu\text{A}$. However, since CV-measurements could be made at high gate voltages, the breakdown was approximated to $8\text{--}10 \text{ MV/cm}$, which is high enough for

MOSFETs (although the leakage itself causes failure). The leakage may in part be caused by high amounts of aluminium in the oxides, which has been seen with Secondary Ion Mass Spectrometry [10].

4. Conclusion

Four different silicon carbide samples have been thermally oxidized. The difference in oxidation rate of the C face and the Si face may be explained by the number of available silicon bonds on respective crystal face. The oxides have been characterized using capacitance-voltage measurements at room temperature. It was found that wet oxidation on n-type results in less oxide charge, but possibly with a higher density of surface states. Oxides on p-type were difficult to characterize, especially on the C face wafers.

Acknowledgements

This work was partly supported by the Swedish National Board for Industrial and Technical Development (NUTEK).

References

1. Bhatnagar, M. and Baliga, B. J., IEEE Trans. Electron Devices **40**, 645 (1993).
2. Davis, R. F., Kelner, G., Shur, M., Palmour, J. W. and Edmond, J.A., Proc. IEEE **79**, 677 (1991).
3. Ivanov, P. A. and Chelnokov, V. E., Semicond. Sci. Technol. **7**, 863 (1992).
4. Muelhoff, L., Choyke, W. J., Bozack, M. J. and Yates, J. T., J. Appl. Phys. **60**, 2842 (1986).
5. Muelhoff, L., Bozack, M. J., Choyke, W. J. and Yates, J. T., J. Appl. Phys. **60**, 2558 (1986).
6. Zheng, Z., Tressler, R. E. and Spear, K. E., J. Electrochem. Soc. **137**, 854 (1990).
7. Kimoto, T., Nishino, H., Yoo, W. S. and Matsunami, H., J. Appl. Phys. **73**, 726 (1993).
8. Singh, N. and Rys, A., J. Appl. Phys. **73**, 1279 (1993).
9. Ouisse, T., Bécourt, N., Jaussaud, C. and Templier, F., J. Appl. Phys. **75**, 604 (1994).
10. Zetterling, C.-M. and Östling, M., Presented at ICSCRM'93, Washington D.C. (1993).
11. Palmour, J. W., Davis, R. F., Kong, H.S., Corcoran, S. F. and Griffis, D. P., J. Electrochem. Soc. **136**, 502 (1989).
12. Massoud, H. Z., Plummer, J. D. and Irene, E. A., J. Electrochem. Soc. **132**, 2685 (1985).
13. Ligensa, J. R., Phys. Chem. **65**, 2011 (1961).
14. Harris, R. C. and Call, R. L., Silicon Carbide 1973, (329, University of South Carolina Press, 1973).
15. Suzuki, A., Ashida, H., Furui, N., Mameno, K. and Matsunami, H., Jpn. J. Appl. Phys. **21**, 579 (1982).
16. Nicollian, E. H. and Brews, J. R., "MOS Physics and Technology", 1st ed. (Wiley, New York, 1982).
17. Petit, J. B., Neudeck, P. G., Matus, L. G. and Powell, J. A., 4th International Conference on Amorphous and Crystalline Silicon Carbide and Other IV-IV Materials, (Springer Proceedings in Physics, Vol 71, 190, Springer-Verlag, 1991).

Strain Relaxation in Epitaxial $\text{Si}_{1-x}\text{Ge}_x$ Layers During some Silicidation Processes

O. Nur, M. R. Sardela Jr., H. H. Radamson, M. Willander, G. V. Hansson and S. Hatzikonstantinidou*

Dept. of Physics and Measurement Technology, Linköping University, S-581 83 Linköping, Sweden.

Received May 9, 1994; accepted June 15, 1994

Abstract

Strain relaxation in $\text{Si}_{1-x}\text{Ge}_x$ epitaxial layers is investigated after the reaction of Co and Pt with strained SiGe alloy. The epitaxial SiGe alloy studied has a Ge fraction ranging between $x = 5\%$ to 22% . The strain in the $\text{Si}_{1-x}\text{Ge}_x$ is evaluated using multi-crystal high resolution x-ray diffraction mapping in reciprocal space (MC-HRXRD). The results show that for Co in order to keep the strain in $\text{Si}_{1-x}\text{Ge}_x$ unaffected, a sacrificial Si layer is needed. The direct reaction of 40 nm Co on $\text{Si}_{0.9}\text{Ge}_{0.1}$ can lead to defect formation and 40% strain relaxation. This is in contrast to Pt/ $\text{Si}_{1-x}\text{Ge}_x$ reaction, where negligible relaxation was observed.

1. Introduction

Silicides have been used both as highly conductive layers and for their rectifying property in different devices, including Schottky diodes, bipolar transistors, metal base transistors, and infrared sensors [1–5]. Moreover, strained $\text{Si}_{1-x}\text{Ge}_x$ epitaxial layers offer the possibility of band-gap engineering according to the used Ge fraction x . This strained alloy, if combined with a silicide, will lead to new heterostructures which is of great technological interest. Properties ranging from metallic to semiconducting state are offered by silicides. An example of this technological interest is the application of the such heterostructure for infra-red detectors. It is well known that infrared radiation is transparent in air in two windows. These are $3\text{--}5 \mu\text{m}$ and $8\text{--}12 \mu\text{m}$ wave lengths. Usually this radiation is detected through a Schottky barrier diode (SBD), which is a barrier between a metal and p-type Si. The silicide – p type silicon has a relatively low Schottky barrier height, so when infrared radiation is shined on such diodes, holes will get enough energy to overcome this barrier and a photon induced current can result from this process. No SBD which can be used to detect infra-red radiation on both windows in the silicide – silicon system was possible to achieve. The Schottky barrier height Φ_p on p type semiconductor is given by the following relation:

$$\Phi_p = q\chi_s + E_g - q\Phi_m \quad (1)$$

Where χ_s and E_g are the electron affinity and energy gap of the semiconductor, respectively, and Φ_m is the work function of the metal. Since the electron affinities of silicon and germanium are very close in magnitude, electron affinity of SiGe is expected to be very close to that of silicon and germanium. So a smaller E_g will lead to a decrease of the barrier height and consequently lead to the possibility of

detecting long wave lengths, which was difficult to achieve in the silicide – silicon structure.

Among the different silicides, CoSi_2 and NiSi_2 are promising candidate for Si-based heterostructure future devices, since they have a cubic lattice structure with small lattice mismatch to silicon and hence the possibility of epitaxial growth on Si. Another important silicide is PtSi , since it has been of interest when combined with p-type SiGe for infrared detectors. The reaction of Pt with strained SiGe alloys has been studied from phase formation point of view but no data has been given about the strain state following the reaction of Pt with strained $\text{Si}_{1-x}\text{Ge}_x$ [6]. In all devices that employ a silicide/strained SiGe alloy the strain should be carefully preserved in order to optimize the device performance, since relaxation of the strain will progress through misfit dislocations that degrade device performance in addition to the fact that the band gap of the partially relaxed SiGe alloy will differ from that of the strained SiGe alloys.

In this paper, we investigate the strain properties of Co/SiGe and Pt/SiGe after thermal treatment using MC-HRXRD reciprocal space mapping. The strain in the SiGe layer is carefully investigated before and after the reaction of Co and Pt with the SiGe alloy.

2. Experimental

On two 3 inch p-Si(001) substrates (samples C1 and C2) SiGe alloy was grown by silicon molecular beam epitaxy (Si-MBE) in a VG V – 80 UHV system having a base pressure of 5×10^{-11} Torr. The MBE system and growth procedures have been described previously [8, 9]. The intended structure has a Si buffer layer and 400 nm $\text{Si}_{0.9}\text{Ge}_{0.1}$ layer. This was followed by the growth of a thin, lightly doped n-Si layer (60 nm) on sample C1. The substrates were then removed from the MBE chamber. This was followed by sputtering 40 nm Co (in sample C1 and C2) and a subsequent rapid thermal process (RTP) anneal in N_2 for 40 seconds. This choice of this annealing regime is due to the fact that it is a standard regime for the CoSi_2 formation. In a third lightly doped n-Si(001) substrate (sample C3), Co was implanted with a dose of $2 \times 10^{17} \text{ cm}^{-2}$ and an energy of 200 keV followed by a 900 °C anneal for 5 minutes. This resulted in an approximately 700 Å thick CoSi_2 buried layer. Then $\text{Si}_{0.9}\text{Ge}_{0.1}$ was grown on top of this buried CoSi_2 layer. Another set of $\text{Si}_{1-x}\text{Ge}_x$ samples with $x = 0.56\%$ – 22.4% (samples C4–C7), were grown in a Balzers UMS 360 MBE system having a base pressure of 5×10^{-11} Torr. The growth procedures and the chamber description can be

* Present address: Dept. of Solid State Electronics, Royal Institute of Technology, Box 1084, 164 21 Kista, Sweden.

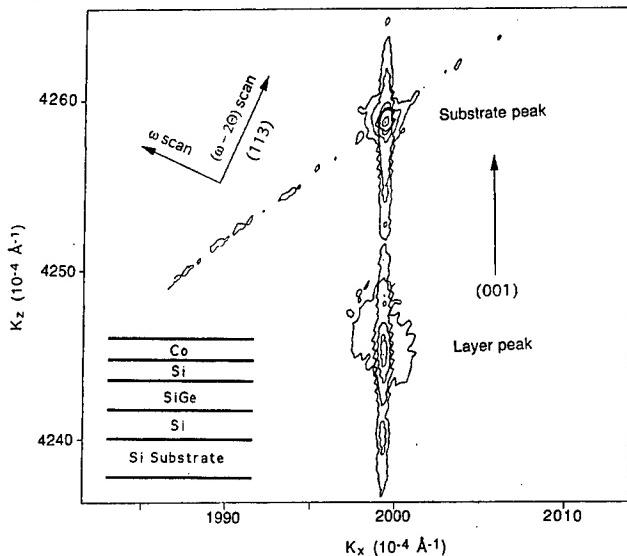


Fig. 1. A 2-D reciprocal space map of a strained $Si_{0.9}Ge_{0.1}$ layer obtained around the [113] diffraction. The $CoSi_2$ was formed from a sacrificial Si layer on top of the strained $Si_{0.9}Ge_{0.1}$ layer (sample D1). The insert is a schematic diagram of the fabricated structure before the RTA stage.

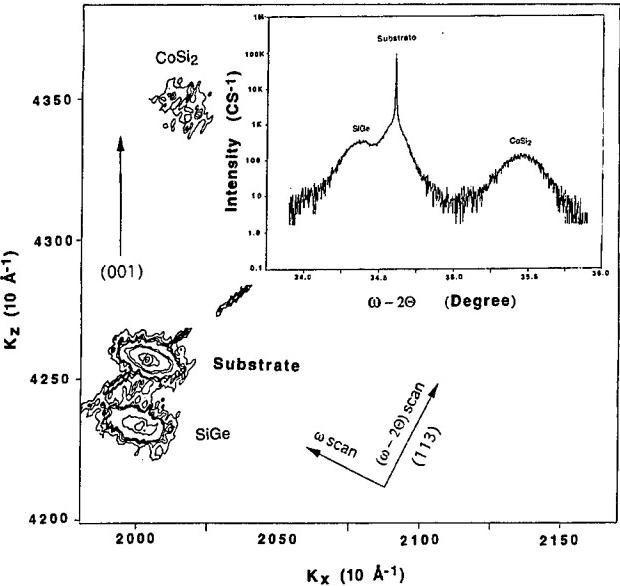


Fig. 3. A 2-D reciprocal space map of a partially relaxed $Si_{0.9}Ge_{0.1}$ layer obtained around the [113] diffraction. The sample contains a buried $CoSi_2$ layer (sample D3). The insert is a high resolution (004) rocking curve showing the Si , $SiGe$, and the $CoSi_2$ peaks.

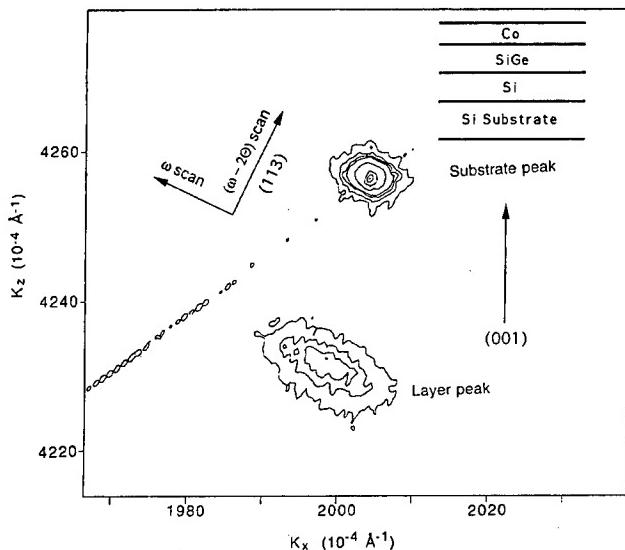


Fig. 2. A 2-D reciprocal space map of a partially relaxed $Si_{0.9}Ge_{0.1}$ layer obtained around the [113] diffraction. The $CoSi_2$ was formed directly on a strained $Si_{0.9}Ge_{0.1}$ layer (sample D2). The insert is a schematic diagram of the fabricated structure before the RTA stage.

found elsewhere [7]. Platinum was sputtered to a thickness of 30 nm on this set of samples and annealed at 350 °C in Ar ambient for 30 minutes. At this annealing temperature, $PtSi$ is expected to be formed.

Strain investigation of the two sets of samples was performed by multi-crystal high resolution X-ray diffraction (MC-HRXRD) mapping. The multi-crystal high resolution diffractometer has a 15 times better instrumental resolution compared to double crystal high resolution diffractometer, and it can detect a relative lattice parameter variation of a layer with respect to the substrate in the range of 10^{-5} . In this work we have used a Philips MRD diffractometer which have a Bartels 4-crystal monochromator using $Ge(220)$ crystals to collimate the Cu-radiation incident beam to 12 arc sec divergence angle. The diffracted beam was also collimated to 12 arc sec by another two crystal $Ge(220)$ col-

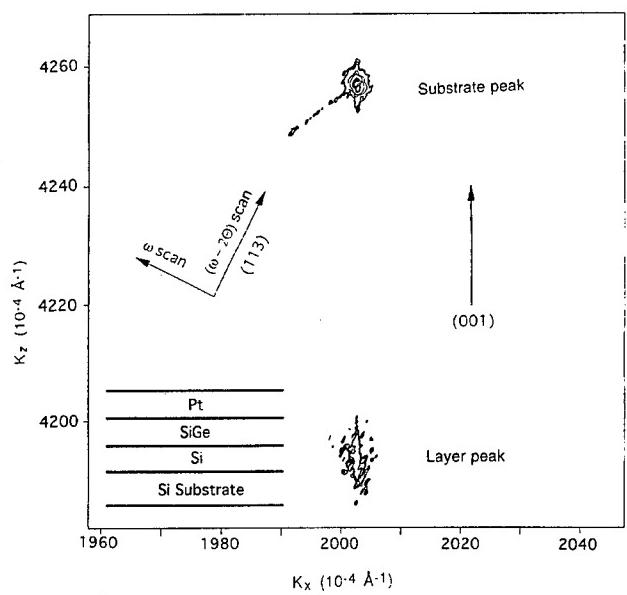


Fig. 4. A 2-D reciprocal space map of strained SiGe (22.4% Ge) layer obtained around the [113] diffraction. The Pt was reacted directly on the SiGe alloy. The insert is the final structure before the thermal treatment.

Table I. Calculated relative lattice parameter variation in the parallel, perpendicular, and the relaxation factor (R) of the SiGe for the as grown and the thermally treated case for the different $Pt/Si_{1-x}Ge_x$ samples

Sample	$f_{par.}$	$f_{per.}$	R
C4 (5.6% Ge)			
As grown	2.6×10^{-5}	4.0×10^{-3}	1.1×10^{-2}
30 nm Pt, 350 °C 30 min.	7.1×10^{-5}	3.8×10^{-3}	1.2×10^{-2}
C5 (10.5% Ge)			
As grown	1.7×10^{-5}	7.3×10^{-3}	4.2×10^{-3}
30 nm Pt, 350 °C 30 min.	7.7×10^{-5}	7.4×10^{-3}	1.8×10^{-2}
C6 (15.5% Ge)			
As grown	5.0×10^{-5}	1.1×10^{-2}	8.3×10^{-3}
30 nm Pt, 350 °C 30 min.	6.5×10^{-5}	1.1×10^{-2}	5.6×10^{-2}
C7 (22.4% Ge)			
As grown	5.0×10^{-5}	1.6×10^{-2}	5.7×10^{-3}
30 nm Pt, 350 °C 30 min.	1.1×10^{-4}	1.5×10^{-2}	2.3×10^{-2}

limator. The most useful pair of angular scans for mapping is a combination of an ω scan and a coupled ω - 2Θ scan (ω and 2Θ are the angles of the sample surface and the diffracted beam, respectively, relative to the incident beam direction). In a reciprocal lattice map, an ω -scan probes the orientation variations for a defined interplanar spacing, while an ω - 2Θ scan probes interplanar spacing variations for the same orientation [12]. By recording the ω and 2Θ values for the layer and substrate peak positions in the map, the relative variation of the layer and substrate lattice constant parallel and perpendicular to the sample surface, the lattice mismatch of a layer with respect to the substrate, in addition to the relaxation factor can be calculated [10, 11]. The advantage of two dimensional reciprocal space mapping as well as the limitation of the multi-crystal X-ray diffractometer can be found elsewhere [12].

3. Results

In order to probe the relative lattice variation of a layer parallel and perpendicular to the growth direction, and hence investigate the strain state of that layer, it is useful to use an asymmetric reflection. We have used for all measurements the (113) reflection which is a convenient reflection from a geometrical point of view. Figure 1 shows the reciprocal space map of sample C1. As it is clear from the alignment of the substrate and layer peaks along the (001) growth direction, the SiGe alloy remains strained after forming the CoSi₂ from the sacrificial Si layer. The calculated relaxation factor for this case was 7×10^{-3} , which is below the 0.02 detection limit. From the shape of the iso-intensity contours it is also clear that no defects have been created in this case since the ratio of the full width at half maxima (FWHM) of the layer and the substrate peaks in the ω scan direction is very close to unity. For the case of the direct sputtering of Co on the SiGe alloy (sample C2), Fig. 2 indicates a partially relaxed alloy with presence of misfit dislocations. The relaxation factor calculated in this case was 0.59 which is about 60% relaxation, and the ratio of the FWHM of the layer and the substrate peaks in the ω scan direction is 10. In the case of the ion implanted buried CoSi₂, Fig. 3 demonstrates the fact that the buried CoSi₂ has a defected top Si layer which lead to a partially relaxed SiGe later with a relaxation factor of 0.4, and a SiGe peak 20 broader at FWHM than the substrate peak in the ω scan direction. From the angular position of the buried CoSi₂ peak, a value of 1.17% lattice mismatch to Si was calculated. This is in consistence with the standard 1.2% mismatch of CoSi₂ to Si. Figure 4 shows the (113) reciprocal

space map of sample C7 (22.4% Ge). As it is clear that the SiGe layer remains strained, since the layer and substrate peaks are vertically aligned along the (001) growth direction. However, the scattered low intensity contours of the layer peak indicates a locally defected region on the SiGe top surface. Table I gives the relaxation factor (R), the relative variation of the layer (SiGe) lattice parameter in the parallel, and the perpendicular directions for all the Pt/Si_{1-x}Ge_x case (C4-C7) as grown and after the reaction of Pt with the SiGe alloy. The small variation of lattice parameter between the as grown structures and the thermally treated case indicates that all the SiGe alloy structures ($x = 5.6\%-22.4\%$) remains essentially strained after the reaction with Pt followed by the 350 °C–30 min. low temperature annealing regime.

4. Conclusion

In conclusion, the strain in Co/Si_{1-x}Ge_x and Pt/Si_{1-x}Ge_x reactions following temperature regimes appropriate for the formation of CoSi₂ and PtSi is investigated. It is clear that the two metals react differently with SiGe alloys. The strain analysis presented here seems to be consistent with previously depth profile, and phase transformation data presented in the literature. It has been shown before, that in the case of Co reaction with Si_{1-x}Ge_x alloys, the reaction yields a CoSi_{1-x}Ge_x compound coexisting with Si_{1-x}Ge_y alloy. While, in the case of Pt, and at this annealing temperature a uniform reacted layer containing a mixture of Pt₂Si and Pt₂Ge is produced.

References

1. Van Der Hove, L., Walters, R., Maex, K., De Keersmaecker, R. F. and Declerck, G. J., IEEE Trans. Electron Dev. **34**, 554 (1987).
2. Akbar, S., Ratanaphanyarat, S., Kuang, J. B., Chu, S. F. and Hsieh, C. M., Electronics Lett. **28**, 86 (1992).
3. Schüppen, A. et al., Electronics Lett. **29**, 215 (1993).
4. Kanaya, H., Hasegawa, F., Yamaka, E., Moriyama, T. and Nakajima, M., Jpn. J. Appl. Phys. **28**, 544 (1989).
5. Liou, H. K. et al., Appl. Phys. Lett. **60**, 577 (1992).
6. Hong, Q. Z. and Mayer, J. W., J. Appl. Phys. **66**, 611 (1989).
7. Sardela, M. R. Jr., Radamsson, H. H., Ni, W.-X., Sundgren, J.-E. and Hansson, G. V., Jpn. J. Appl. Phys. **33**, (1994) 417.
8. Hasan, M. A. et al., J. Vac. Sci. Technol. **B5**, 1332 (1987).
9. Knall, J., Sundgren, J.-E., Hansson, G. V. and Greene, J. E., Surface Sci. **166**, 512 (1986).
10. van der Sluis, P., Jpn. J. Appl. Phys. **26**, 188 (1993).
11. Hornstra, J. and Bartels, W. J., J. Cryst. Growth **44**, 513 (1984).
12. Fewster, P. F., Semicond. Sci. Technol. **8**, 1915 (1993).

An Investigation of the Stability of Copper Germanide Thin Films in the Presence of Si and SiO_2

J. P. Doyle, B. G. Svensson, M. O. Aboelfotoh* and J. Hudner

Royal Institute of Technology, Solid State Electronics, P.O. Box E229, S-164 40 Kista-Stockholm, Sweden

Received June 13, 1994; accepted June 15, 1994

Abstract

Copper germanide (Cu_3Ge) has received interest in recent years as a potential metallization for VLSI applications due to its exceptionally low room temperature resistivity. We have investigated the thermal stability of Cu_3Ge thin films on both silicon and thermally oxidized silicon wafers. Films were deposited by electron beam evaporation of sequential layers of Ge and Cu and exposed to an annealing schedule ranging from 100°C to 450°C. Secondary ion mass spectrometry (SIMS) analysis has revealed an interaction of the film with the silicon substrate. At temperatures as low as 200°C, diffusion of silicon into the copper germanide film was observed with the concentration and depth of penetration scaling with increased annealing temperature. Results on controlling this interaction will also be presented and correlation is made with resistivity measurements.

1. Introduction

Recent reports in the literature have described a low resistivity compound of Cu_3Ge and proposed it as a possible candidate for use as an interconnect or contact metallization in silicon-based VLSI technology [1, 2]. Although numerous investigations involving copper-silicon reactions have been conducted [3–8], far fewer have been carried out in the copper-germanium system [5, 8]. Additionally we are aware of only one systematic study of the interaction of silicon with copper germanium, however, this was in the presence of excess copper [8]. In this study, we have examined the temperature stability of stoichiometric layers of Cu_3Ge in the presence of silicon to determine the resistance of the films to silicon penetration in the temperature range of 100°C to 450°C.

2. Experimental

Cu_3Ge thin films were formed via e-beam evaporation of sequential layers of Cu and Ge deposited onto substrates held at room temperature at a rate of 0.4 nm/sec. The substrates consisted of silicon (100) wafers and thermally oxidized silicon wafers. All substrates were solvent cleaned prior to loading into the vacuum chamber with the silicon substrates receiving a diluted HF strip as well. The system base pressure was less than 3×10^{-7} Torr and during the evaporation was less than 1×10^{-6} Torr. Stoichiometric films were formed from elemental films whose thicknesses were calculated assuming bulk densities. The thickness of the films was 240 nm and in order to assess the reactivity with the substrate, the multilayer films were fabricated with

both copper and germanium in direct contact with the silicon substrate. Anneals were performed in either vacuum or in a tube furnace continuously purged with nitrogen. The annealing time in all cases was 30 minutes.

Material characterization of the films was performed by Secondary Ion Mass Spectrometry (SIMS) utilizing a Cameca IMS 4f microanalyzer. A primary sputtering beam of 5.5 keV O_2^+ ions was rastered over an area with a size of $200 \times 200 \mu\text{m}^2$, and positive secondary ions were collected from the central region (diameter approximately 60 μm) of the sputtered crater. Sheet resistance measurements were made using a standard four point probe. Film thicknesses were determined using a Tencor Alpha Step 200 surface stylus profilometer.

3. Results

Figures 1 and 2 show the profiles from copper germanide films formed at 100°C and 150°C for 30 minutes, respectively. Here the germanium layer was in direct contact with the silicon prior to the anneal. In both cases the observed silicon signal in the bulk of the film is more than four orders of magnitude less than the copper and germanium signals, indicative of the resistance of the films to the penetration of silicon. The uniformity of the copper and germanium signals after a 150°C anneal is consistent with the formation of Cu_3Ge as this is the only phase known to form in this temperature regime [5]. In Fig. 3, a 150°C annealed film where copper was in direct contact with the substrate is shown with no evidence of silicon penetration observed.

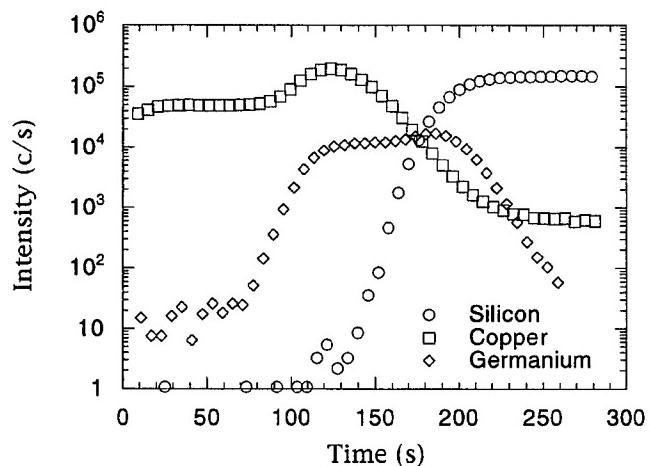


Fig. 1. SIMS profiles of copper and germanium layers after a 100°C anneal for 30 minutes. The non-uniformity in the copper and germanium signals is representative of incomplete reaction of the layers. The layering sequence had germanium in direct contact with the silicon.

* Present address: North Carolina State University, Department of Materials Science and Engineering, Raleigh, North Carolina 27695-7907, U.S.A.

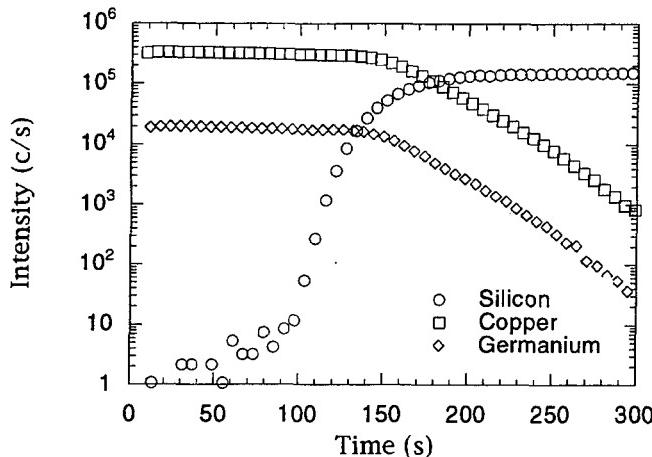


Fig. 2. SIMS profiles after a 150 °C anneal. The uniform copper and germanium signals are indicative of the completed reaction of the layers to form Cu₃Ge. The layering sequence had germanium in direct contact with the silicon.

Results from sheet resistance measurements as a function of annealing temperature are shown in Figs 4 and 5 for both copper and germanium in contact with the substrate. Regardless of the stacking arrangement, a plateau in the sheet resistance is observed for temperatures in excess of

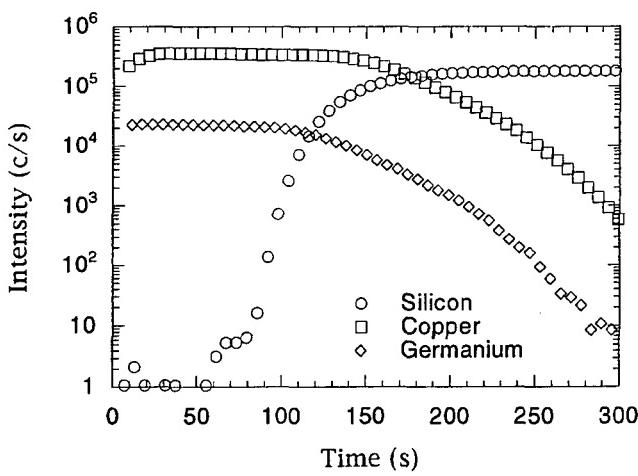


Fig. 3. SIMS profiles after a 150 °C anneal with the layering sequence having copper in contact with the silicon.

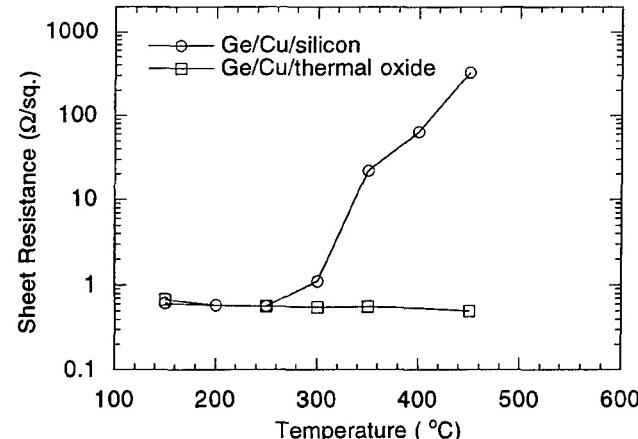


Fig. 4. Sheet resistance measurements on films formed from the layering sequence with copper in direct contact with either silicon or thermally formed silicon dioxide.

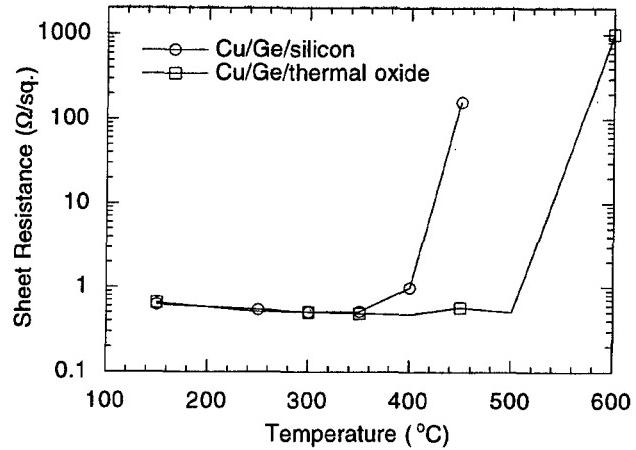


Fig. 5. Sheet resistance measurements on films formed from the layering sequence with germanium in contact with either silicon or thermally formed silicon dioxide. Note that the point at 600 °C is scaled down by a factor of 1000.

150 °C which agrees favorably with the SIMS results for Cu₃Ge formation; however, an increase in resistivity is observed to occur in the range of 250–300 °C for films where copper was deposited directly onto the silicon. Films with germanium in contact with the silicon displayed an increase in resistivity in a temperature range between 350–400 °C. In the case of films deposited on thermal oxide, the resistivity remained constant to temperatures on the order of 450 °C irrespective of the initial stacking arrangement. To determine the upper limit of temperature stability, two films on thermal oxide were annealed at higher temperatures. At 500 °C, the film begins to display a cloudy surface, although no corresponding change in resistivity was observed. At 600 °C, however, the film displayed a dull, frosted appearance, and a sheet resistance beyond the measurement capability of our four point probe (>10⁶ ohm/sq.) indicative, presumably of agglomeration and a discontinuous film. Agglomeration at this temperature is not unexpected, as the melting point of Cu₃Ge is reported to be 749 °C [8].

SIMS profiling was then undertaken on samples annealed in the range of 100 to 400 °C to determine the silicon content of the films. As shown in Fig. 6, for films with copper initially in contact with the substrate, a two orders of magnitude increase in the silicon signal is observed after a

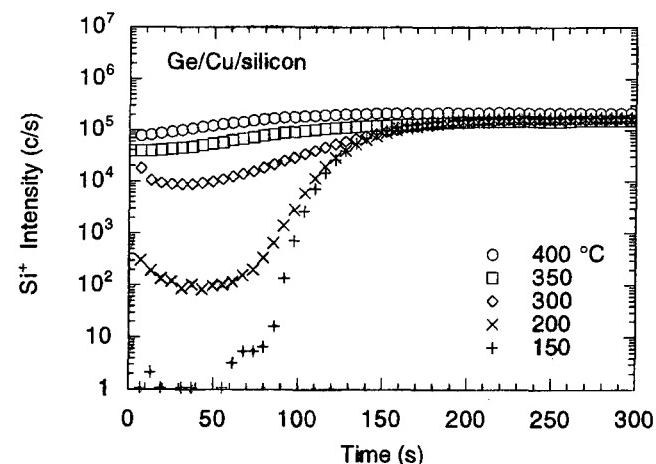


Fig. 6. SIMS profile of the silicon signal for films annealed at various temperatures on films where the layering sequence had copper in contact with the silicon substrate.

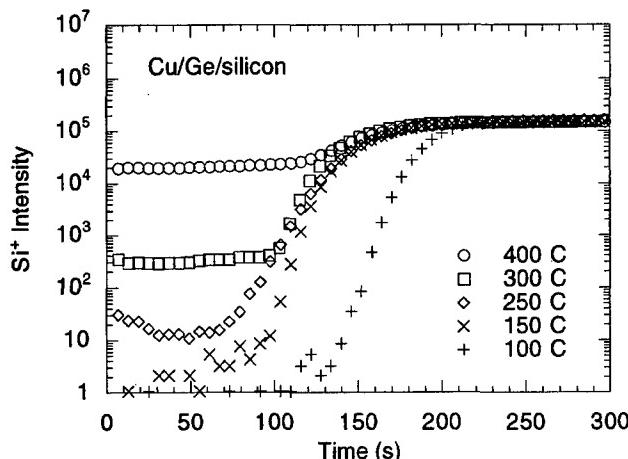


Fig. 7. SIMS profile of the silicon signal for films annealed at various temperatures where the layering sequence had germanium in contact with the silicon substrate.

200 °C anneal as compared to the 150 °C anneal. In contrast, a film with germanium in contact with the silicon displays the same silicon level between 250 °C and 300 °C anneal as shown in Fig. 7.

At higher temperatures, the level of silicon in the film rises rapidly, regardless of the initial ordering of the layers, with the silicon signal equalling or even surpassing the germanium level.

4. Discussion

From Figs 6 and 7, it is apparent that the temperature at which silicon incorporation into the Cu₃Ge begins depends on the initial stacking arrangement of the elemental films. With copper in direct contact, silicon is observed in the film at a temperature of 200 °C, whereas, when germanium is in direct contact, the same level of silicon occurs between 250 °C and 300 °C. Although the reaction of copper and germanium is completed by 150 °C (Fig. 2), earlier SIMS studies on the formation of copper silicide, have shown that temperatures in excess of 150 °C are required to form Cu₃Si and a temperature of 200 °C is necessary to completely react a 100 nm copper layer with silicon [9]. The lack of silicon in our films annealed at or below 150 °C is consistent with this observation. On the other hand, XPS studies have reported the formation of an interfacial layer of a metal rich Cu₃Si-like phase during evaporation of copper at room temperature [10]. This phase may act as a seed layer for the growth of Cu₃Si resulting in a lower temperature penetration of silicon into the Cu₃Ge film via the formation of Cu₃Si. In the absence of this phase, a higher formation temperature may be required as in the case where germanium is in contact with the silicon. Independent of this, the fact that in both cases, silicon penetration occurs at or near the formation temperature of copper silicide, suggests the formation of Cu₃Si at the expense of Cu₃Ge. The observation by d'Heurle and Gupta of free germanium at the surface of a Cu₃Ge film in the presence of silicon is consistent with this

view although the lack of thermodynamical data makes it impossible to confirm [8].

The sheet resistance of the Cu₃Ge films, however, is not sensitive to small amounts of silicon incorporation. Not until the silicon content is within about 10% of the germanium level, does a rise in the resistivity appear, an indication that if Cu₃Si forms, it is composed of small inclusions of Cu₃Si interspersed in the Cu₃Ge film. The resistivity of Cu₃Si, on the order of 100 microhm-cm, is approximately one order of magnitude greater than that of Cu₃Ge. In the case of thermal oxide, the sheet resistance values are found to be constant in the range of 150 °C to 500 °C, indicating that the incorporation of silicon is not as severe as the silicon substrates. Preliminary SIMS analysis of these films does not reveal any significant silicon at the front surface independent of the stacking arrangement.

5. Conclusions

Cu₃Ge can be formed on both silicon and thermally formed silicon dioxide at temperatures as low as 150 °C; however, silicon incorporation is found in the Cu₃Ge films at temperatures as low as 200 °C, when deposited on silicon substrates. The temperatures at which these levels of silicon are incorporated are relatively low when compared to current silicon device processing temperatures. Increases in resistivity are directly correlated to silicon content. The penetration of silicon into the films may be the result of the formation of Cu₃Si at the expense of Cu₃Ge an indication of the higher stability of the silicide. On thermally formed silicon dioxide, Cu₃Ge ceases to be electrically stable in the range of 500 to 600 °C.

Acknowledgements

Fruitful discussions with Professor F. M. d'Heurle are gratefully appreciated. Financial support was received from the Swedish Board for Technical Development (NUTEK).

References

1. Aboelfotoh, M. O., Tawancy, H. M. and Krusin-Elbaum, L., *Appl. Phys. Lett.* **63**, 1622 (1993).
2. Aboelfotoh, M. O. and Krusin-Elbaum, L., *J. Appl. Phys.* **70**, 3382 (1991).
3. Becht, J. G. M., van Loo, F. J. J. and Metelaar, R., *Reactivity of Solids* **6**, 3655 (1988).
4. Veer, F. A., Kolster, B. H. and Burgers, W. G., *Trans. Met. Soc. AIME* **242**, 669 (1968).
5. Hong, S. Q., Comrie, C. M., Russell, S. W. and Mayer, J. W., *J. Appl. Phys.* **70**, 3655 (1991).
6. Stolt, L., d'Heurle, F. M. and Harper, J. M. E., *Thin Solid Films* **200**, 147 (1991).
7. Becht, J. G. M., van Loo, F. J. J. and Metelaar, R., *Reactivity of Solids* **6**, 61 (1988).
8. d'Heurle, F. M. and Gupta, J., *Appl. Surf. Sci.* **73**, 214 (1993).
9. Svensson, B. G., Aboelfotoh, M. O. and Lindstrom, J. L., *Phys. Rev. Lett.* **66**, 3028 (1991).
10. Cros, A., Aboelfotoh, M. O. and Tu, K. N., *J. Appl. Phys.* **67**, 3328 (1990).

Properties of Iron Silicide Contacts to n- and p-type Silicon

U. Erlesand and M. Östling*

Royal Institute of Technology, Solid State Electronics Electrum 229, S-164 40 Kista – Stockholm, Sweden

Received June 2, 1994; accepted in revised form June 15, 1994

Abstract

Thin polycrystalline films of FeSi and β -FeSi₂ were formed by solid state reaction with either boron or phosphorus doped silicon. The corresponding electrical heterojunction properties were studied by temperature activated current–voltage and capacitance–voltage analysis. The FeSi/Si Schottky characteristics showed that the current transport mechanism was dominated by thermionic emission across the interfaces for both n- and p-type silicon substrates. Indications of recombination through deep levels was only detected for n-type silicon measured by current–voltage at low temperatures. The Schottky barrier heights of FeSi were estimated to 0.68 ± 0.03 eV and 0.40 ± 0.03 eV at 0 K for respectively n- and p-type silicon. The Schottky barrier was observed to be pinned to the silicon valence band. The formation of β -FeSi₂ produced current transport characteristics with ideality factors of about 1.02–1.05 on both n- and p-type silicon indicating no recombination through deep levels inside the silicon depletion layer. The CV results on p-type silicon strongly showed the presence of shallow defects or neutral complex formation.

1. Introduction

The semiconducting phase of iron disilicide, β -FeSi₂, has been shown to be a suitable material for a future light emitting silicon device [1]. The band structure indicates a direct transition of about 0.85–0.89 eV [2–6], and therefore considered ideal as a material in an opto-coupler for fibre communication. To date only a few publications [6–10] have been reported on the electrical transport properties across iron-silicide–silicon junctions. Epitaxial β -FeSi₂/n-Si and β -FeSi₂/p-Si diodes investigated by Lefki *et al.* [6] showed non-rectifying behaviour due to large non-ideal leakage currents. The epitaxial β -FeSi₂/n-Si diodes studied by Regolini *et al.* [7], formed by selective chemical vapour deposition, showed rectifying behaviour but with diode ideality factors as large as 3. Regolini also studied β -FeSi₂/p-Si diodes which showed ohmic behaviour. Radermacher *et al.* [8] studied the electrical behaviour of ion beam synthesised buried metallic α -FeSi₂/n-Si diodes and obtained ideality factors of 1.4. Dimitriadi [9] investigated polycrystalline β -FeSi₂/n-Si diodes formed by solid state reaction at 850 °C in H₂ atmosphere and concluded that current was dominated by multistep tunnelling via deep levels. Early investigations of iron deposited on clean vacuum cleaved silicon showed IV ideality factors of 1.17–1.70 [11]. These investigations all show the presence of non-ideal currents across the silicide–silicon junction. Since iron is known to have a high impurity diffusivity [12] and to act as a deep level impurity in both n and p-type silicon [13], the question to ask is whether the presence of iron at elevated temperature induces these defects or if they may be related to the fabrication method used. High quality beta iron silicide contacts to n-type

silicon are crucial for the fabrication of the light emitting device where electrons are injected from silicon to silicide, therefore making it necessary to evaluate how iron affects current transport in n-type silicon. The close proximity of n- and p-type regions in modern VLSI techniques also makes it vital to investigate the iron influences on p-type silicon.

In this investigation the electrical transport properties were studied for the metallic iron monosilicide (FeSi) and the transport ideality for the semiconducting (β -FeSi₂) heterojunctions to n- and p-type silicon.

2. Experimental procedure

n- and p-type Czochralski silicon (111) substrates were used in this investigation. The p-type samples were boron doped to a resistivity of about $7.6\Omega\text{ cm}$ while the n-type samples had an $30\mu\text{m}$ thick phosphorus doped $9\Omega\text{ cm}$ epitaxial layer on Sb doped ($10^{-3}\Omega\text{ cm}$) bulk wafers. The samples were cleaned according to the following procedure: 5 min in hot 130 °C sulphuric acid, 5 min in 80 °C nitric acid, a 30 s dilute HF (HF : H₂O, 1 : 10) dip and subsequently oxidized in a wet atmosphere forming about 800 nm silicon dioxide. Standard lithographic procedure was used to produce circular openings with diameters ranging from 15 to 1000 μm . The samples were then cleaned using hot 130 °C sulphuric and 80 °C nitric acids with a final dip in dilute HF (HF : H₂O, 1 : 10) for 30 seconds and a 5 min water rinse.

The samples were immediately loaded into an electron beam-heated evaporator equipped with turbomolecular and ion pumps and evacuated to approximately 4×10^{-8} Torr. The pressure increased to 10^{-7} Torr during evaporation. A 33 nm iron and a 5 nm silicon cap layer were sequentially deposited. The iron and silicon deposition rates were respectively 0.5 and 1 nm/s.

The samples were annealed in a vacuum furnace in the low 10^{-7} Torr range for one hour at 450 °C to form FeSi, for one hour at 650 °C or 800 °C to form β -FeSi₂. Unreacted iron on the oxide areas was removed by a 5 sec dip in buffered hydrofluoric acid (BHF) followed by wet etching in an sulphuric acid mixture for 3 min followed by a final dip in BHF for 5 seconds. The samples annealed at 650–800 °C were then immediately loaded into an electron-beam evaporator for contact metallization and evacuated to 10^{-6} Torr. About 300 nm of aluminium was evaporated at a rate of 3 nm/s. The samples were then finally patterned and wet etched. The Al contact reduce probe contact resistance and distribute applied voltage homogeneously to the disilicide. It does not itself influence the IV and CV measurements [10] at the used voltages and temperatures. Rutherford back-scattering spectrometry (RBS) was carried out, using 2.4 MeV ${}^4\text{He}^+$ ions with a sample tilt of 60° with respect to

Present address: Stanford University, CIS, Stanford, CA 94305, USA

normal incidence, to estimate silicide stoichiometry and thickness. X-ray diffraction (XRD) using CuK α radiation was also used to verify the formed silicide phases. Current voltage characteristics was measured between 77–300 K in a liquid nitrogen cryostat using an HP4145B IV-analyser. Schottky barrier height values for the iron monosilicide were determined by extrapolating the forward current-voltage characteristics to zero applied voltage. The effective Richardson constants were assumed to be 112 and 32 A/K 2 for respectively electrons and holes [14]. The slope of the linear portion of the log-current voltage curve yielded the value for the ideality factor for both the metal–silicon and the semiconductor–silicon systems, see discussion below. In those cases where a linear fit could not be made over at least two orders of magnitude because of series resistance the extrapolation methods of Lien *et al.* [15] were applied. Capacitance voltage measurements were evaluated in the temperature and voltage range of respectively 77–300 K and 0–5 V using an HP 1 MHz CV-meter.

3. Results and discussion

3.1. Silicide formation

Figure 1 shows the results of RBS and XRD analysis from samples annealed at 450 and 650 °C. Annealing at 800 °C produce similar RBS results at 650 °C. RBS stoichiometry determination display the formation of iron monosilicide (FeSi) at 450 °C and iron disilicide (β -FeSi₂) at 650–800 °C. Using the densities of 6.17 and 4.93 g/cm 3 from Nicolet *et al.* [16] for FeSi and β -FeSi₂ respectively, the analysis showed a formation of 64 nm FeSi and 111 nm β -FeSi₂. The formed phases were confirmed by XRD which also displayed the polycrystalline nature of the films.

3.2. Electrical evaluation

The forward logarithmic current voltage (I -V) characteristics as a function of temperature for samples annealed at 450 °C are displayed in Fig. 2. The ideality factor n for temperatures between 295 and 195 K shows a value below 1.05 independent of temperature. Since fairly high mobility materials are used, $\mu > 1000 \text{ cm}^2/\text{Vs}$, the ideality factor indicates that the dominant current transport mechanism is due to thermionic emission ($I_{\text{thermionic}}$) [17]. No evident voltage dependence of the ideality factor is seen in the linear

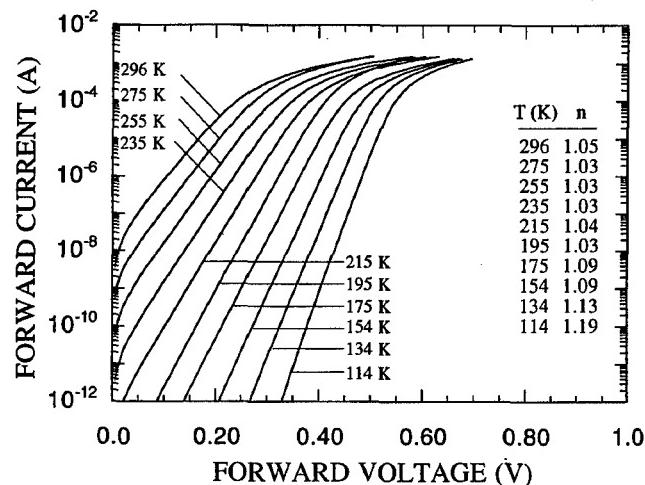


Fig. 2. Forward current–voltage characteristics of the FeSi/n-Si junction as a function of temperature for silicides formed at 450 °C. The diode area was $3.14 \times 10^{-4} \text{ cm}^2$.

portions of the curves, a fact that is more clearly observed by evaluating a logarithmic plot of $I/(1 - \exp(-qV/kT))$ against V [18]. This result together with the temperature independent ideality factor indicate that the current flow due to recombination via defect levels in the depletion region (I_{rec}) must be small. The deviation of n from unity may be due to the image-force lowering or from interface effects [18]. Below 195 K the ideality factor is seen to increase with decreasing temperature, which is probably due to the reduction of the thermionic emission current causing an increased $I_{\text{rec}}/I_{\text{thermionic}}$ ratio. Thermionic field emission and field emission are ruled out because of the low doping level in the substrates, $5 \times 10^{14} \text{ cm}^{-3}$, and because of the used temperature interval [18]. The temperature IV activation plot for thermionic emission behaviour, I_0/T^2 vs. $1/T$, shown in Fig. 3, results in a curve with two distinct slopes, i.e. activation energies. The slope of $0.68 \pm 0.03 \text{ eV}$, in the temperature independent region of n , yields the barrier height at 0 K according to the ideal thermionic emission model [19].

The temperature dependence below 195 K is described by an activation energy of $0.52 \pm 0.03 \text{ eV}$. This value, approximately half the bandgap of Si, is presumably associated with the recombination current which causes deviation from ideal thermionic emission behaviour at low temperatures

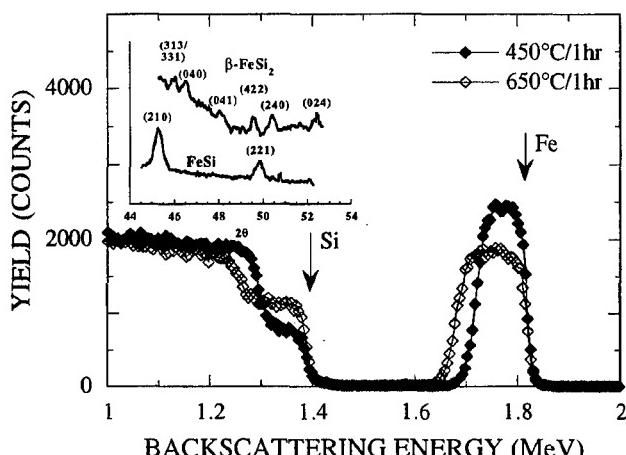


Fig. 1. RBS spectra showing the formation of fully reacted and stoichiometric FeSi and β -FeSi₂ after annealing at 450 °C and 650 °C respectively. The inset shows the corresponding XRD spectra.

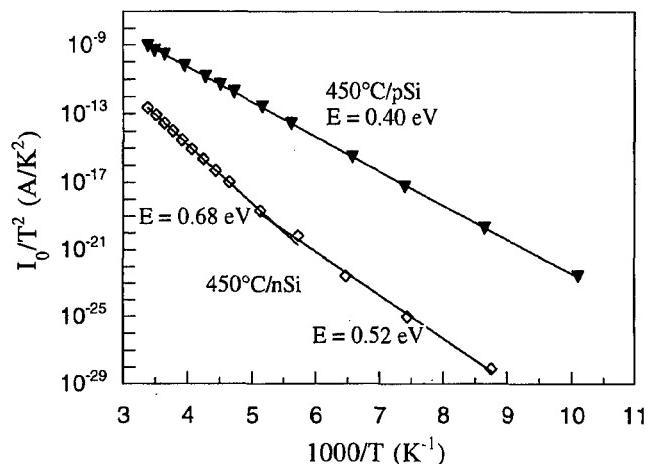


Fig. 3. Temperature activation IV plot of the forward leakage current I_0 for both n and p-type samples.

[20]. The barrier height values determined from *IV* measurements are therefore only reliable at temperatures above 195 K. This is observed in Fig. 4. The Schottky barrier heights determined by *IV* is slowly increasing with decreasing temperature to 195 K. Below 195 K the SBH drops which indicates the failure of the thermionic emission model [21–22]. This is also supported by the *CV* analysis. The *CV* measurements, shown in Fig. 5 for 295 and 77 K, display ideal characteristics over the applied temperature and voltage range. No influence of deep levels were observed. The SBH temperature dependence obtained both by *CV* and *IV* show a value of approximately 10^{-4} eV/K, approximately displaying the total silicon bandgap temperature change. The forward *IV* characteristics for samples annealed at 450 °C for 60 min forming FeSi on boron doped (111) silicon are plotted in Fig. 6. The samples display ideality factors around 1.06 independent of the temperature in the range 95 to 295 K. No voltage dependence other than series resistance could be observed. The temperature activation *IV* plot, shown in Fig. 3, indicates a single dominant current transport mechanism. The Schottky barrier height was extracted to 0.40 ± 0.03 eV at 0 K. The results indicate that thermionic emission dominates the current transport across the FeSi/p-type silicon interface. The Schottky barrier heights as determined from *IV* measurements are shown in

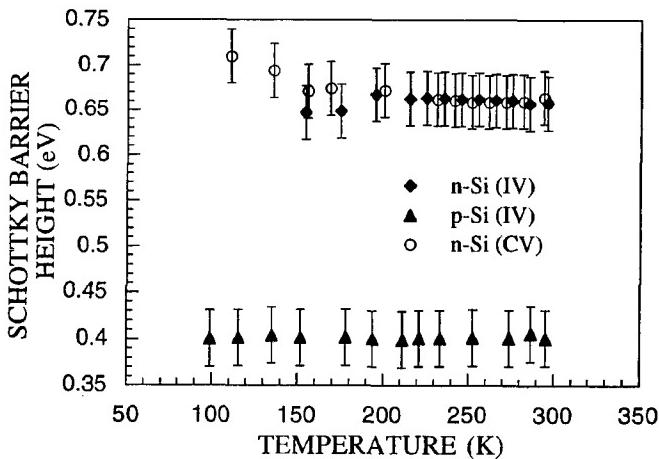


Fig. 4. Temperature dependence of the FeSi Schottky barrier heights on n- and p-type silicon.

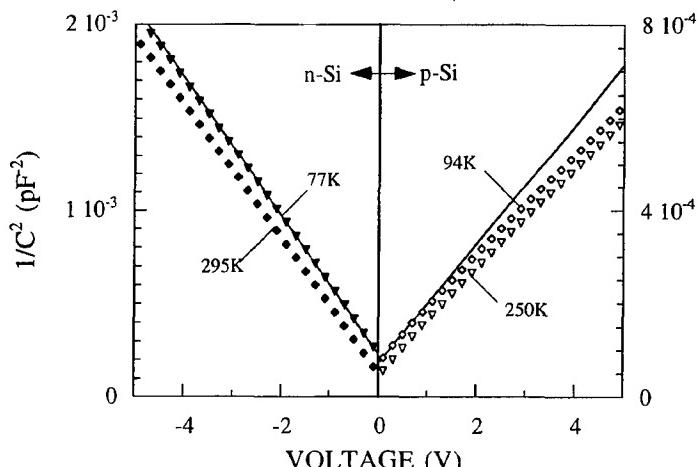


Fig. 5. Capacitance–voltage relations for FeSi/Si diodes at some selected temperatures. The diode area was $8.4 \times 10^{-3} \text{ cm}^2$.

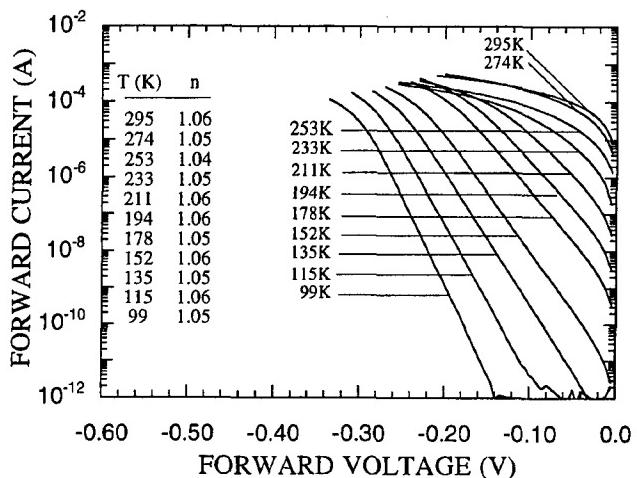


Fig. 6. Forward current–voltage characteristics of the FeSi/p-Si junction as a function of temperature for silicides formed at 450 °C. The diode area was $2 \times 10^{-4} \text{ cm}^2$.

Fig. 4. The obtained temperature coefficient of the barrier height is very small. This together with the temperature dependence on n-type silicon therefore indicates that the FeSi barrier height on silicon is pinned to the valence band [23–24]. The sum of the measured Schottky barrier heights obtained by *IV* for n- and p-type silicon at 295 K gives a value of about 1.07 ± 0.06 eV, i.e. a deviation of less than 10% from the bandgap of silicon, and indicate that the barrier is only pinned to the valence band. All *CV* measurements on p-type samples (450–800 °C) show a substantial curvature in the $1/C^2$ vs. *V* plot, as seen in Fig. 5 which displays analysis at 250 and 94 K for a silicide formed at 450 °C. This behaviour excluded any barrier height or diffusion potential evaluation from the *CV* measurements. The curvature (about the same order for all formation temperatures) indicates a reduced effective dopant concentration near the interface in the order of $5\text{--}8 \times 10^{14} \text{ cm}^{-3}$. Experiments on boron redistribution during the formation of iron silicides have shown only smaller dopant depletion close (within 1000 Å) to the silicide contact [25] and can therefore not be used to explain the observed effect. The temperature independent reduction indicate that the curvature in the $1/C^2$ vs. *V* plot is probably due to carrier trapping at a shallow defect level ($T > 77 \text{ K} \rightarrow E_t - E_v > 0.15 \text{ eV}$) or due to dopant passivation by formation of some neutral iron–boron related complexes. To the authors knowledge no neutral iron–boron related complexes has been reported to affect the boron doping profiles in silicon, but complexes can form according to the iron–boron phase diagram. The electrically active iron–boron acceptor ($E_c - E_t \approx 0.1 \text{ eV}$) is also shown to completely dissolve at temperatures higher than 400 °C [26]. The cause of the observed effect is therefore not known. The measured *IV* ideality factor show no signs of deep levels inside the depletion layer. The observed anomaly in the *CV* study may therefore be caused by (a) an electrically neutral iron–dopant complex passivation or (b) an electrical defect level at $E_t - E_v \leq 0.15 \text{ eV}$ with a reduced concentration closer to the silicide interface not affecting the forward current transport.

The forward current–voltage characteristics obtained after annealing samples at 650 °C or 800 °C for 1 hr [10] display very similar behaviour as the Schottky diodes. At these temperatures, according to XRD and RBS, the Fe film

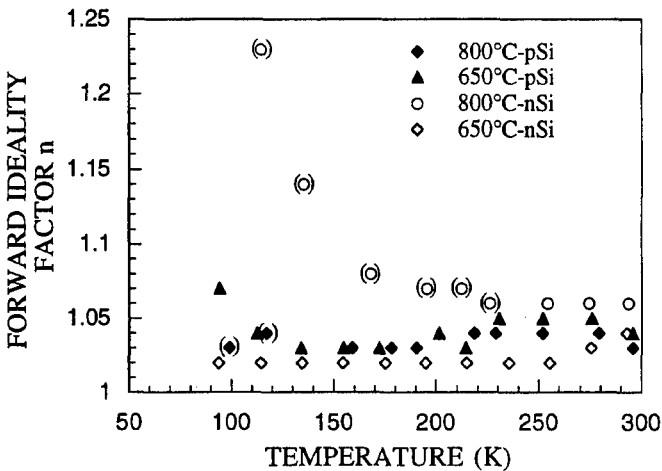


Fig. 7. Forward current-voltage ideality factors for the $\beta\text{-FeSi}_2/\text{Si}$ junctions as a function of temperature. Parenthesis () marks observed ideality factor voltage dependency.

has been completely transformed into the semiconducting $\beta\text{-FeSi}_2$ phase, producing a semiconductor-semiconductor anisotype heterojunction. The reported intrinsic doping levels in the as formed $\beta\text{-FeSi}_2$ films, p-type $7\text{-}12 \times 10^{18} \text{ cm}^{-3}$ [5, 7] are much larger than in the silicon substrate and the relative dielectric constant for the silicide is about twice the value of silicon [3, 4]. The diffusion potential and the amount of applied voltage drop on the $\beta\text{-FeSi}_2$ side of the heterojunction is then much smaller than on the silicon side. Using the resistivity of $9\Omega\text{cm}$ (a carrier concentration of $5 \times 10^{14} \text{ cm}^{-3}$ as measured by CV) for the n-type silicon substrate one obtains the relative band bending and depletion layer widths to approximately $V_b(\text{silicide})/V_b(\text{Si}) \approx 2 \times 10^{-4}$ respectively $W(\text{silicide})/W(\text{Si}) \approx 5 \times 10^{-4}$. It is therefore a good approximation to assume that all voltage drop appears within the silicon. The current can then be expressed as

$$I \approx I_0 \exp \left[\frac{qV}{nkT} \right] \quad (1)$$

where I_0 is independent of applied voltage if it is determined by either diffusion [27, 28], thermionic emission [18] or recombination via interface states [29, 30]. Recombination via deep levels [31] may also produce similar current behaviour but with a voltage dependence in the pre-exponential

factor. The measured IV curves, displayed in reference [10], for 650°C samples display low, voltage and temperature independent ideality factors over the range 95 to 295K , as shown in Fig. 7. This shows that the recombination current through deep defects levels inside silicon is small. The influence of recombination is only observable at low temperatures for the samples formed at 800°C . Evaluation of the temperature activation energy requires some further investigation since the true transport mechanism is unknown. The capacitance-voltage relation for the heterojunctions can be simplified to the following expression if in a first approximation the interface states are neglected [27] and recalling that the product of dielectric constant and carrier concentration in the silicide is much larger than in the silicon substrate.

$$\frac{1}{C^2} \approx \frac{2}{(qN_d \epsilon)} \left[V_b - 2 \frac{kT}{q} - V \right] \quad (2)$$

where V_b is the diffusion potential or band bending inside the silicon region, N_d and ϵ are respectively the doping level and the dielectric constant in the silicon substrate. The ideal shape of the $1/C^2$ vs. V plot therefore display a linear relationship. Fig. 8 shows the characteristics for silicides formed at 650 and 800°C . The 650°C -sample indicates a small nonlinearity of the curve with the apparent effect of an increased active carrier concentration towards the bulk. The anomaly may be caused by deep defect levels or by dopant redistribution. Previous experiments on dopant redistribution [25] have indicated only small disturbances close to the interface and is therefore ruled out. In the figure is also shown the measurement of the capacitance at 82K , displaying a slightly reduced nonlinearity. Formation of deep defect levels as the cause of this effect is therefore likely. Preliminary results have indicated the formation of a deep level at $E_c\text{-}0.3\text{eV}$ [10]. The CV characteristics of 800°C samples, show a more ideal behaviour indicating a reduced concentration of defects during reverse biasing. Comparing the results from IV and CV for the 650 and 800°C samples shows that the amount of defects formed far from the interface at 650°C are reduced when annealed at 800°C , but the number of defects closer to the interface has increased.

4. Conclusions

Thin polycrystalline films of FeSi and $\beta\text{-FeSi}_2$ have been formed by solid state reaction with boron or phosphorus doped silicon and the corresponding electrical heterojunction properties studied. The Schottky barrier heights produced by FeSi were estimated to $0.68 \pm 0.03\text{eV}$ and $0.40 \pm 0.03\text{eV}$ at 0K for respectively n- and p-type silicon. The current transport mechanism across the FeSi/Si junction is dominated by thermionic emission over a large temperature interval with ideality factors lower than 1.06. The Schottky barrier was observed to be pinned to the silicon valence band. The transport mechanism across the $\beta\text{-FeSi}_2/\text{n-Si}$ and p-Si junctions formed at 650°C show no signs of recombination via deep levels inside the depletion layer. The CV results on p-type silicon strongly indicated the presence of shallow defects or neutral complex formation while the IV results showed no indications of deep level formation. Similar results were obtained for n-type silicon but with much less influence of deep defect levels. The corre-

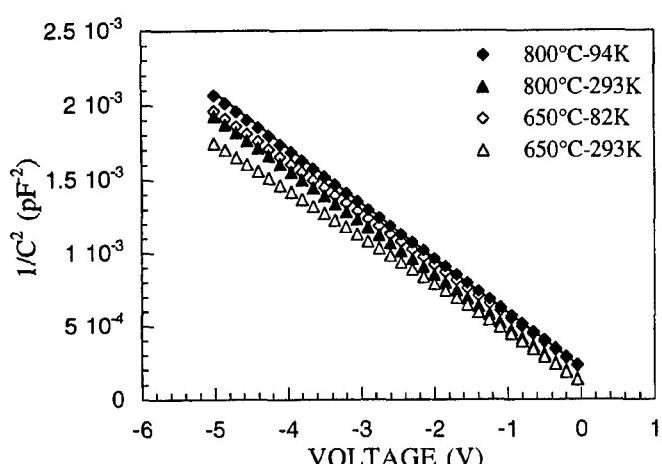


Fig. 8. Capacitance-voltage relations for $\beta\text{-FeSi}_2/\text{Si}$ diodes at some selected temperatures. The diode area was $8.4 \times 10^{-3} \text{ cm}^2$.

lation of IV and CV analysis then indicate that the electrically active defect levels are depleted in concentration towards the silicide silicon interface.

Acknowledgements

The authors are very grateful for interesting and fruitful discussions with Dr B. G. Svensson. This project was supported financially by the Swedish Research Council for Engineering Sciences (TFR).

References

1. Bost, M. C. and Mahan, J. E., *J. Vac. Sci. Technol.* **B4**, 1336 (1986).
2. Birkholz, U. and Shelm, J., *Phys. Status. Sol.* **34**, K177 (1969).
3. Bost, M. C. and Mahan, J. E., *J. Appl. Phys.* **58**, 2696 (1985).
4. Bost, M. C. and Mahan, J. E., *J. Appl. Phys.* **64**, 2034 (1988).
5. Dimitriadis, C. A., Werner, J. H., Logothidis, S., Stutzmann, M., Weber, J. and Nesper, R., *J. Appl. Phys.* **68**, 1726 (1990).
6. Lefki, K., Muret, P., Cherif, N. and Cinti, R. C., *J. Appl. Phys.* **69**, 352 (1991).
7. Regolini, J. L., Trincat, F., Berbezier, I. and Shapira, Y., *Appl. Phys. Lett.* **60**, 956 (1992).
8. Radermacher, K., Mantl, S., Dieker, Ch. and Lüth, H., *Appl. Phys. Lett.* **59**, 2145 (1991).
9. Dimitriadis, C. A., *J. Appl. Phys.* **70**, 5423 (1991).
10. Erlesand, U., "Semiconducting Iron Disilicide on Silicon" (Thesis Royal Inst. of Technology, Stockholm, Sweden).
11. van Otterloo, J. D. and Gerritsen, L. J., *J. Appl. Phys.* **49**, 723 (1978).
12. Weber, E. R., "Properties of Silicon" (INSPEC, London 1987), p. 421.
13. Chen, J.-W. and Milnes, A. G., *Ann. Rev. Mater. Sci.* **10**, 157 (1980).
14. Andrews, J. M. and Lepseter, M. P., *Solid-St. Electron.* **13**, 1011 (1970).
15. Lien, C.-D., So, F. C. T. and Nicolet, M.-A., *IEEE Trans. Electron. Dev.* **31**, 1502 (1984).
16. Nicolet, M.-A. and Lau, S. S., "VLSI Electronics Microstructure Science" (Edited by N. G. Einspruch and G. B. Larrabee) (Academic Press, New York 1983), Vol. 6.
17. Crowell, C. R. and Sze, S. M., *Solid-St. Electron.* **9**, 1035 (1966).
18. Rhoderick, E. H. and Williams, R. M., "Metal-Semiconductor Contacts" (Clarendon, Oxford 1988), 2nd ed.
19. Bethe, H. A., MIT Radiation Lab. Rep. 43-12 (1942).
20. Yu, A. and Snow, E., *J. Appl. Phys.* **39**, 3008 (1968).
21. Aboelfotoh, M. O. and Tu, K. N., *Phys. Rev.* **B33**, 6572 (1986).
22. Aboelfotoh, M. O., *J. Appl. Phys.* **61**, 2558 (1987).
23. Aboelfotoh, M. O. and Tu, K. N., *Phys. Rev.* **B34**, 2311 (1986).
24. Aboelfotoh, M. O., *J. Appl. Phys.* **64**, 4046 (1988).
25. Erlesand, U. and Östling, M., *Appl. Surf. Sci.* **73**, 186 (1993).
26. Gao, X., Mollenkopf H. and Yee, S., *Appl. Phys. Lett.* **59**, 2133 (1991).
27. Anderson, R. L., *Solid-St. Electron.* **5**, 341 (1962).
28. Shokley, W., *Bell Syst. Tech. J.* **28**, 435 (1949).
29. Fonash, S. J., "Solar cell device physics" (Academic Press, New York 1981).
30. Barnett, A. M. and Rothwarf A., *IEEE Trans. Electron Dev.* **27**, 615 (1980).
31. Sah, C. T., Noyce, R. N. and Schokley, W., *Proc. IRE* **45**, 1228 (1957).

List of Participants

Veli-Matti Airaksinin, Univ. of Technology, Otakaari 7A, FIN-02150 Espoo, Finland	Francois D'Heurle, KTH Electrum 229, Dept. of Electronics, S-16440 Kista, Sweden	Hafliði P. Gíslason, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland	Sofia Hatzikonstantinidou, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden
Ylva Bäcklund, Univ. of Uppsala, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	J. P. Doyle, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Magnus Gisselfällt, Chalmers University, Dept. of Applied Physics, S-41296 Göteborg, Sweden	Isak Sverrir Hauksson, Heriot-Watt University, Dept. of Physics, Edinburgh EH14 4AS, Scotland
Mietek Bakowski, IMC, P.O. Box 1084, S-16421 Kista, Sweden	Bengt Edholm, Uppsala University, Dept. of Technology, S-75121 Uppsala, Sweden	Jaime Gordon, Odense University, Fysik Institut, DK-5230 Odense, Denmark	Päivi Heimala, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland
Karl-Fredrik Berggren, Linköping University, Physics and Meas. Technology, S-58183 Linköping, Sweden	Tryggvi Egilsson, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland	Kaj Grahn, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Detlef Heitmann, Inst. f. Angewandte Physik, Jungiusstr. 11, D-20355 Hamburg, Germany
Mats Bergh, Chalmers University, Solid State Electronics, S-41296 Göteborg, Sweden	Pontus Eriksson, KTH, Sorögatan 45, NB, S-16441 Kista, Sweden	Jostein Grepsstad, NTH Dept. of Physical Electron., N-7034 Trondheim, Norway	Jouni Heleskivi, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland
Jan Borglind, IMC, Isafjordsgatan 22, S-16421 Kista, Sweden	Ulf Erlesand, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Viktoras Gruzhinskis, Semiconductor Physics Institute, Gostauto 11, 2600 Vilnius, Lithuania	Peter Hessling, Chalmers University, Dept. of Applied Physics, S-41296 Göteborg, Sweden
Arne Bratas, Norw. Institute of Technology, Division of Physics, N-7034 Trondheim, Norway	Marco Fanciulli, Aarhus University, Inst. of Physics and Astronomy, DK-8000 Aarhus, Denmark	Halldór F. Guðjónsson, Science Institute, University of Iceland, Dunhagi 3, 107 Reykjavik, Iceland	Anssi Hovinen, University of Technology, Otakaari 7A, FIN-02150 Espoo, Finland
Markus Büttiker, IBM Research Center, Box 218, Yorktown Heights, NY 10598, U.S.A.	Tor A. Fjeldly, Norw. Institute of Technology, Dept. of Physical Electronics, N-7034 Trondheim, Norway	Vidar Guðmundsson, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland	Ola Hunderi, NTH, Dept. of Physics, N-7034 Trondheim, Norway
Juan Cardenas, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Sami Fransilla, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Christer Hallin, IFM-FOA, Linköping University, S-58183 Linköping, Sweden	Jørn Hvam, DTH, Bldg. 345 East, DK-2800 Lyngby, Denmark
K. A. Chao, Univ. of Trondheim, Inst. of Physics, NTH, N-7034 Trondheim, Norway	Christer Fröjd,br/> KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Michiel Hamberg, MESA Research Institute, P.O. Box 217, NL-7500 AE Enschede, Netherlands	Olle Inganäs, IFM, Link. University, Lab. of Applied Physics, S-58183 Linköping, Sweden
Thomas Clausen, Technical University, Physics Dept., Bldg. 309, DK-2800 Lyngby, Denmark	Yuri Galperin, University of Oslo, Physics Department, P.O. Box 1048, Blindern, N-0316 Oslo, Norway	Anders Hanneborg, University of Oslo, P.O. Box 124, Blindern, N-0314 Oslo, Norway	Snorri Ingvarsson, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland
Gordon Davies, Kings College, Physics Department, Strand, London WC2R 2LS, England	Thomas Garm, Univ. of Aalborg, Pontoppidanstr. 103, DK-9220 Aalborg, Denmark	Erland Brun Hansen, J.C. Orsted Institute, Universitetsparken 5, DK-2100 Copenhagen, Denmark	Mats Isberg, Uppsala University, Dept. of Technology, S-75121 Uppsala, Sweden

Erik Janzén, IFM-FOA, Linköping University, S-58183 Linköping, Sweden	Kari Leinonen, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Ferenc Masszi, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Ritta Paananen, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland
Mats Jonson, Chalmers University, Dept. of Appl. Physics, S-41296 Göteborg, Sweden	Kristján Leósson, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavík, Iceland	Algirdas Matulis, Semiconductor Physics Institute, Gostauto 11, 2600 Vilnius, Lithuania	Chris Palmstrøm, Bellcore, 331 Newman Spring Rd, 3X-201 Red Bank, NJ 07701, U.S.A.
Per G. Jonsson, Uppsala University, Dept. of Technology, S-75121 Uppsala, Sweden	Seppe Leppävuori, University of Oulu, Microelectronics Lab., Linnanmaa, FIN-90570 Oulu, Finland	Jyrki Molarius, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	André Paulsen, Norw. Telecom Research, Box 83, N-2007 Kjeller, Norway
Bert Junno, Lund University, Solid State Physics, Box 118, S-22100 Lund, Sweden	Eva Lind, University of Uppsala, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Bo Monemar, IFM, Linköping University, S-58183 Linköping, Sweden	Kjeld Pedersen, Aalborg University, Pontoppidanstr. 103, DK-9220 Aalborg, Denmark
Tord E. Karlén, KTH Electrum 229, Dept. of Electronics, S-16440 Kista, Sweden	Karin Ljungberg, University of Uppsala, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Per Morgen, Univ. Odense, Campusvej 55, DK-5230 Odense, Denmark	Sture Petersson, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden
Jyrki Kiihamäki, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Ágústa Loftsdóttir, University of Iceland, Dunhagi 3, IS-107 Reykjavík, Iceland	Hans-Erik Nilsson, KTH Electrum 229 Solid State Electronics, S-16440 Kista, Sweden	Jón Pétursson Science Institute, Univ. of Iceland, Dunhagi 3, IS-107 Reykjavík, Iceland
Klaus v. Klitzing, MPI f. Festkörperforschung, Heisenbergstr. 1, D-70569 Stuttgart, Germany	Stian Lovold, Defence Research Establishm., Box 25, N-2007 Kjeller, Norway	Kenneth Nordgren, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Joakim Petersson, Chalmers University, Dept. of Physics, Fysikgränd 2, S-41296 Göteborg, Sweden
Bernhard Kowalski, Lund University, Solid State Physics, Box 118, S-22100 Lund, Sweden	Nils Lundberg, KTH Electrum 229, Dept. of Electronics, S-16440 Kista, Sweden	Edvard Nordlander, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Gernot Pomrenke, AFOSR/NE Bldg. 410, 110 Duncan Ave, suite B115, Bolling AFB, D.C., U.S.A.
Sigurgeir Kristjánsson, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavík, Iceland	Lennart Lundqvist, IMC, Box 1084, S-16421 Kista, Sweden	Kai Nordlund, Helsinki Univ. of Technology, P.O. Box 43, FIN-00014 Helsinki, Finland	Risto Punkkinen, University of Turku, Lemminkäisenkatu 14A, FIN-20520 Turku, Finland
Pekka Kuivalainen, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Jari Mäkinen, University of Helsinki, Otakaari 1 M, FIN-02150 Espoo-Helsinki, Finland	Omar Nur Linköping University, Dept. of Physics, S-58183 Linköping, Sweden	H. H. Radamson Linköping University, Dept. of Physics, S-58183 Linköping, Sweden
Ulf E. Hvam Laheld, Norwegian Institute of Technology, Dept. of Physics, N-703 Trondheim, Norway	Michael Markow, European Office of Aerospace Research, 223/231 Old Marylebone Rd, London NW1 5TH, England	Ulla Övergaard, University of Uppsala, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Lars-Åke Ragnarsson, Chalmers University, Solid State Electronics, S-41286 Göteborg, Sweden
Arne N. Larsen, University of Aarhus, Inst. of Physics and Astronom., DK-8000 Aarhus C, Denmark	Jörgen Olsson, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Sven-Ingmar Ragnarsson, Nutek, S-16421 Kista, Sweden	

Lennart Ramberg, IMC, Box 1084, S-16421 Kista, Sweden	Ulf Sannemo, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Einar Ö. Sveinbjörsson, Chalmers University, Solid State Electronics, S-41296 Göteborg, Sweden	Zhaogua Wang, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland
Tuomo Rantala, University of Oulu, Microelectronics Lab., Linnanmaa FIN-90570 Oulu, Finland	Gerd Schuppner, KTH Electrum 229, ESD-Lab., S-16440 Kista, Sweden	Kalman Tarnay, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Gerlinde Xander, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland
Pelle Rangsten, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Bo Sernelius, Linköping University, Dept. of Physics and Measurement Technology, S-58183 Linköping, Sweden	Peer Tidemand-Petersson, Tele Danmark Research, Lyngsø Allé 2, DK-2970 Horsholm, Denmark	Hongqi Xu, University of Lund, Solid State Physics, S-22100 Lund, Sweden
Mats Rosling, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	R. I. Shekhter, Chalmers University, Dept. of Applied Physics, S-41296 Göteborg, Sweden	Göran Thungström, KTH Electrum 229, Solid State Electronics, S-16440 Kista, Sweden	Baohua Yang, Science Institute, University of Iceland, Dunhagi 3, IS-107 Reykjavik, Iceland
Jaakko Saarilahti, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Juha Sinkkonen, Helsinki University of Technology, Otakaari 7A, FIN-02150 Espoo, Finland	Stefan Tiensuu, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Trond Ytterdal, Norwegian Institute of Technology, Dept. of Physical Electronics, N-7034 Trondheim, Norway
Jorma Salmi, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Konstantin Smekalin, University of Technology, Dept. of Physics, Box 692, FIN-33101 Tampere, Finland	Olof Tornblad, KTH Electrum 229, Dept. of Electronics, S-16440 Kista, Sweden	Alexandre Zagoskin, Chalmers University, Dept. of Applied Physics, S-41296 Göteborg, Sweden
Jaakko Salonen, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Anders Söderbärg, Uppsala University, Dept. of Technology, Box 534, S-75121 Uppsala, Sweden	Bruno Ullrich, University of Tokyo, Dept. of Physics, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan	Carl-Mikael Zetterling, KTH Electrum 229, Dept. of Electronics, S-16440 Kista, Sweden
Hans Sigg, Paul Scherrer Institut, Badenerstr. 569, CH-8048 Zürich, Switzerland	Henrik Stubb, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland	Shumin Wang, Chalmers University, Dept. of Physics, S-41296 Göteborg, Sweden	Shi-Li Zhang, KTH Electrum 229, FTE, S-16440 Kista, Sweden
Lars Samuelson, Lund University, Solid State Physics, Box 118, S-22100 Lund, Sweden	Iikka Suni, VTT Electronics, Box 11012, FIN-02044 Espoo, Finland		

Physica Scripta

Manuscripts

Shall be sent in triplicate – one original and two copies to any of the editors or to the editorial office.

Editorial Office

Physica Scripta
The Royal Swedish Academy of Sciences
Box 50005, S-104 05 Stockholm, Sweden

A. Bárány, Physica Scripta

University of Stockholm, Atomic Physics, Frescativägen 24,
S-104 05 Stockholm, Sweden
Fax +46-8-158674, Email BARANY@MSI.SUNET.SE

R. Wäppling, Physica Scripta, Department of Physics,

University of Uppsala, Box 530, S-751 21 Uppsala, Sweden
Fax +46-18-183524, Email ROGERW@FYSIK.UU.SE

Editors

General and cross-disciplinary physics

P. L. Christiansen, Physica Scripta, Laboratory of Applied Mathematical Physics, The Technical University of Denmark, Bldg 303, DK-2800 Lyngby, Denmark
Fax +45-4591235, Email LG@LAMF.DTH.DK

M. Høgh Jensen, Physica Scripta, Niels Bohr Institute and Nordita, Blegdamsvej 17, DK-2100 København Ø, Denmark
Fax +45-31389157, Email MHJENSEN@NBIVAX.NBI.DK

K. Mørk, Physica Scripta, Institute of Physics
University of Trondheim/AVH, N-7055 Dragvoll, Norway
Fax +47-73-591852, Email KJELL.MORK@AVH.UNIT.NO

P. Pierański, Physica Scripta, IFM PAN, Smoluchowskiego 17, 60-179 Poznań, Poland
Fax +48-61-684524,
Email PIOTRP@MARTA.IFMPAN.POZ.EDU.PL

High energy physics

K. Enqvist, Physica Scripta, Research Institute for Theoretical Physics, University of Helsinki, PO Box 9, FIN-00014 Helsinki, Finland
Email ENQVIST@PHCU.HELSINKI.FI

E. Lillestøl, Physica Scripta, PPE-Div, CERN, CH-1211 Geneve 23, Switzerland
Fax +41-22-7820168,
Email EGIL_LILLESTOL@MACMAIL.CERN.CH

U. Lindström, Physica Scripta, Department of Theoretical Physics, University of Stockholm, Box 6730, S-113 85 Stockholm, Sweden
Fax +46-8-347817, Email UL@VANA.PHYSTO.SE

Nuclear physics

G. B. Hagemann, Physica Scripta, Niels Bohr Institute, Tandem Accelerator Laboratory, DK-4000 Roskilde, Denmark
Fax +45-42373516, Email GH@NBITAL.NBI.DK

D.-O. Riska, Physica Scripta, Department of Physics, University of Helsinki, Box 9, FIN-00014 Helsinki, Finland
Fax +358-0-1918378, Email RISKA@PHCU.HELSINKI.FI

Atomic, molecular and optical physics

Dž. Belkić, Physica Scripta, University of Stockholm, Atomic Physics, Frescativägen 24, S-104 05 Stockholm, Sweden
Fax +46-8-158674, Email BELKIC@MSI.SUNET.SE

L. J. Curtis, Physica Scripta, Department of Physics & Astronomy, University of Toledo, Toledo, OH 43606, USA
Fax +1-419-5372723, Email LJC@UTPHYA.PHYA.UTOLEDO.EDU

J. Javanainen, Physica Scripta, Department of Physics, University of Connecticut, Storrs, CT 06269-3046, USA
Fax +1-203-4863346, Email JAVANAI@UCONNVM

I. Martinson, Physica Scripta, Department of Physics, University of Lund, Sölvegatan 14, S-223 62 Lund, Sweden
Fax +46-46-104709, Email INDREK.MARTINSON@FYSIK.LU.SE

A. Rosén, Physica Scripta, Department of Physics, Chalmers University of Technology, S-412 96 Göteborg, Sweden
Fax +46-31-165176, Email F3CAR@FY.CHALMERS.SE

V. P. Shevelko, Physica Scripta, P.N. Lebedev Physical Institute, Leninsky pr. 53, 117924 Moscow, Russia

Fax +7-095-135-2408, Email OSPEC@SPEC.FIAN.MSK.SU

S. Svanberg, Physica Scripta, Division of Atomic Physics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden
Fax +46-46-104250

G. Werth, Physica Scripta, Institut für Physik, Universität Mainz, Postfach 3980, D-55099 Mainz, Germany
Fax +49-6131-395169

Plasma physics

H. L. Pécseli, Physica Scripta, Department of Physics, The University of Oslo, PO Box 1048, Blindern, N-0316 Oslo 3, Norway
Fax +47-22-855671, Email HANS.PECSELI@FYS.UIO.NO

L. Stenflo, Physica Scripta, Department of Plasma Physics, Umeå University, S-901 87 Umeå, Sweden
Fax +46-90-166673

M. Y. Yu, Physica Scripta, Theoretische Physik I, Ruhr Universität, D-44780 Bochum, Germany
Fax +49-234-7094201,
Email YU@PHOENIX.TP1.RUHR-UNI-BOCHUM.DE

Condensed matter physics and material sciences

P. Apell, Physica Scripta, Department of Applied Physics, Chalmers University of Technology, S-412 96 Göteborg, Sweden
Fax +46-31-416984, Email APELL@FY.CHALMERS.SE

H. P. Gislason, Physica Scripta, Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland
Fax +354-1-28911, Email HAFLIDI@RAUNVIS.HI.IS

H. G. Grimmeiss, Physica Scripta, Department of Solid State Physics, Lund Institute of Technology, Box 118, S-221 00 Lund, Sweden
Fax +46-46-104709, Email FTFHG@MACPOST.LU.SE

V. Guðmundsson, Physica Scripta, Science Institute, University of Iceland, Dunhaga 3, IS-107 Reykjavik, Iceland
Fax +354-1-28911, Email VIDAR@RAUNVIS.HI.IS

P. C. Hemmer, Physica Scripta, Department of Physics, NTH, University of Trondheim, N-7034 Trondheim, Norway
Email HEMMER@PHYS.UNIT.NO

T. Jøssang, Physica Scripta, Department of Physics, University of Oslo, Box 1048, Blindern, N-0316 Oslo, Norway
Fax +47-22-855101, Email JOSSANG@FYS.UIO.NO

P. E. Lindelof, Physica Scripta, Niels Bohr Institute for Astronomy, Physics and Geophysics, University of Copenhagen, H C Ørsted Institute, Universitetsparken 5, DK-2100 København Ø, Denmark

R. Nieminen, Physica Scripta, Laboratory of Physics, Helsinki University of Technology, FIN-02150 Espoo, Finland
Fax +358-0-4513116, Email RNIEMINE@CSC.FI

T. Riste, Physica Scripta, Institute for Energy Technology, Postboks 40, N-2007 Kjeller, Norway
Fax +47-63-810920

W. R. Salaneck, Physica Scripta, IFM, University of Linköping, S-581 83 Linköping, Sweden
Fax +46-13-137568, Email WRS@IFM.LIU.SE

T. I. Sigfusson, Physica Scripta, University of Iceland, Science Institute, Dunhaga 3, IS-107 Reykjavik, Iceland
Fax +354-1-28911, Email THIS@RAUNVIS.HI.IS

Geophysics, astronomy and astrophysics

S. J. Johnsen, Physica Scripta, Science Institute, Dunhaga 3, IS-107 Reykjavik, Iceland
Fax +354-1-28911, Email SIGFUS@RAUNVIS.HI.IS

J. M. Knudsen, Physica Scripta, Physics Laboratory, H C Ørsted Institute, Universitetsparken 5, DK-2100 København Ø, Denmark
Fax +45-35320460

Topical Issues in Physica Scripta

Year of Publication	Title	Price per copy in Swedish kronor postage included
1982	2nd General Conference of the Condensed Matter Division of the European Physical Society. Eds. V. Heine, V. R. Van Doren and J. T. Devreeze. 150 pages.	110
	T2: I&II 1982 International Conference of Plasma Physics. Ed. H. Wilhelmsson. 600 pages.	450
1983	Production and Physics of Highly Charged Ions. Ed. L. Liljeby. 260 pages.	195
	T4 Nordic Conference on Surface Science. Eds. M. Pessa and R. Nieminen. 215 pages.	160
	T5 4th Nordic Meeting on Nuclear Physics. Eds. J. Bondorf and G. Hagemann. 230 pages.	170
	T6 Fourth International Workshop on Inelastic Ion-Surface Collisions. Ed. P. Sigmund. 184 pages	140
1984	European Workshop on Very Hot Astrophysical Plasmas. Eds. L. Koch-Miramond and T. Montmerle. 244 pages	185
	T8 Colloquium on Atomic Spectra and Oscillator Strengths for Astrophysics and Fusion Research. Ed. U. Litzén. 152 pages	115
	T9 The Physics of Chaos and Related Problems. Ed. S. Lundqvist. 224 pages. (Nobel Symposium 59)	170
	T10 Disordered Systems by J. Hertz. 42 pages.	30
1985	T11 Astrophysical Aspects of the Interstellar Medium and Star Formation. Ed. G. Gahm. 84 pages. (Crafoord Symposium).	65
	T12 Quantum Fields and Laser Spectroscopy. Ed. S. Stenholm. 76 pages.	60
1986	T13 6th General Conference of the Condensed Matter Division of the EPS. Ed. G. Grimvall. 324 pages.	245
	T14 Fundamental Effects in Semiconductors. Ed. P. Landsberg. 102 pages.	75
	T15 Unification of Fundamental Interactions. Eds. L. Brink, R. Marnelius, J. S. Nilsson, P. Salomonson and B.-S. Skagerstam. 212 pages. (Nobel Symposium 67).	160
1987	T16 The Role of Alpha Particles in Magnetically Confined Fusion Plasmas. Eds. M. Lisak and H. Wilhelmsson. 178 pages.	135
	T17 8th International Conference on Vacuum Ultraviolet Radiation Physics. Eds. P. O. Nilsson and J. Nordgren. 250 pages.	190
	T18 6th International Symposium on Solar Terrestrial Physics. Ed. B. Hultqvist. 316 pages.	235
	T19A + B 7th General Conference of the Condensed Matter Division of the EPS. Eds. F. Bassani, G. Grossi, G. Pastori-Parravicini and M. P. Tosi. 628 pages.	470
1988	T20 An Introduction to Nonlinear Dynamics and Chaos Theory by J. L. McCauley. 60 pages.	45
	T21 Vacuum in Non-Relativistic Matter-Radiation Systems. Eds. F. Persico and E. A. Power. 128 pages.	95
	T22 Workshop and Symposium on the Physics of Low-energy Stored and Trapped Particles. Eds. A. Bárány, A. Kerek, M. Larsson, S. Mannervik and L.-O. Norlin. 330 pages.	250
	T23 Trends in Physics EPS-7. Eds. T. Åberg and S. Stenholm. 336 pages.	250
	T24 Soft Magnetic Materials 8. Eds. H. Pfützner and R. Wäppling. 74 pages.	55
	T25 8th General Conference of the Condensed Matter Division of the EPS. Eds. F. Beleznay, J. Kollár, I. Kovács, N. Kroo and N. Menyhárd. 376 pages.	280
1989	T26 20th EGAS Conference of the European Group for Atomic Spectroscopy. Ed. L. Windholz. 100 pages	75
	T27 Physics of Low-dimensional Systems. Eds. S. Lundqvist and N. R. Nilsson. 168 pages. (Nobel Symposium 73).	125
	T28 Carbon and Oxygen Collision Data for Fusion Plasma Research. Ed. R. K. Janev. 112 pages.	85
	T29 9th General Conference of the Condensed Matter Division of the EPS. Eds. J. Friedel, J. P. Laheurte and J. P. Romagon. 308 pages.	230
1990	T30 Large Amplitude Waves and Fields in Plasmas. Eds. R. Bingham, U. de Angelis, P. K. Shukla and L. Stenflo. 232 pages.	175
	T31 9th International Conference on Vacuum Ultraviolet Radiation Physics. Eds. D. A. Shirley and G. Margaritondo. 312 pages.	235
	T32 6th Nordic Meeting on Nuclear Physics. Eds. G. Lövhöiden, T. F. Thorsteinsen and J. S. Vaagen. 240 pages.	185
	T33 Third Nordic Symposium on Computer Simulation in Physics, Chemistry, Biology and Mathematics. Eds. K. Kaski and M. Salomaa. 248 pages.	185
	T34 22nd EGAS Conference of the European Group for Atomic Spectroscopy. Eds. A. Amesen and R. Hallin. 112 pages.	85
1991	T35 10th General Conference of the Condensed Matter Division of the EPS. Eds. A. G. Vallera, J. B. Sousa, A. F. Martins and L. Alcácer. 312 pages.	235
	T36 The Birth and Early Evolution of Our Universe. Eds. J. S. Nilsson, B. Gustafsson and B.-S. Skagerstam. 304 pages. (Nobel Symposium 79)	230
	T37 Collision Processes of Metallic Ions in Fusion Plasmas. Ed. R. K. Janev. 120 pages.	90
	T38 Fourth Nordic Symposium on Computer Simulations in Natural Sciences. Eds. A. Hansen, E. L. Hinrichsen and P. A. Slotte. 124 pages.	95

Topical Issues in Physica Scripta

Year of Publication	Title	Price per copy in Swedish kronor postage included
T39	11th General Conference of the Condensed Matter Division of the EPS. Eds. J. L. Beeby, P. A. Maksym and J. M. McCoy. 400 pages.	300
1992		
T40	23rd EGAS Conference of the European Group for Atomic Spectroscopy. Ed. S. Łęgowski. 80 pages.	60
T41	The 2nd International Workshop on Auger Spectroscopy and Electronic Structure (IWASES-II). Eds. K. Wandelt, C.-O. Almlöf and R. Nyholm. 304 pages.	230
T42	Low Dimensional Properties of Solids. Eds. M. Jonson and T. Claeson. 220 pages. (Nobel Jubilee Symposium)	165
T43	Extragalactic Astronomy including Observational Cosmology. Ed. A. Elvius. 70 pages. (Craford Symposium)	55
T44	Contributions to Problems in Statistical Physics Elasticity and Dislocation Theory. Eds. T. Jøssang and D. M. Barnett. 160 pages	120
T45	12th General Conference of the Condensed Matter Division of the EPS. Eds. V. Smid, B. Velický and J. Kristofik. 320 pages.	240
1993		
T46	Heavy Ion Spectroscopy and QED Effects in Atomic Systems. Eds. I. Lindgren, I. Martinson and R. Schuch. 272 pages. (Nobel Symposium 85)	205
T47	4th International Colloquium on Atomic Spectra and Oscillator Strengths for Astrophysical and Laboratory Plasmas. Eds. D. S. Leckrone and J. Sugar. 208 pages.	155
T48	Quantum Phase and Phase Dependent Measurements. Eds. W. P. Schleich and S. M. Barnett. 144 pages.	110
T49A+B	13th General Conference of the Condensed Matter Division of the EPS in conjunction Arbeitskreis Festkörperphysik of the DPG. Eds. H. Hoffmann, R. Klein and M. Schwoerer. 752 pages.	565
1994		
T50	Wave-Particle Interaction and Energization in Plasmas. Eds. P. K. Shukla, U. de Angelis, R. Bingham and L. Stenflo. 128 pages.	190
T51	The 25th EGAS Conference of the European Group for Atomic Spectroscopy. Eds. D. Lecler and J. Margerie. 84 pages.	126
T52	Acceleration and Radiation Generation in Space and Laboratory Plasmas. Eds. R. Bingham, J. M. Dawson, T. Katsouleas and L. Stenflo. 160 pages.	240
T53	8th International Seminar on Electron and Ion Swarms. Eds. T. H. Lövaas and H. R. Skulander. 88 pages.	130
T54	16th Nordic Semiconductor Meeting. Eds. H. P. Gíslason and V. Guðmundsson. 312 pages.	468
T55	14th General Conference of the Condensed Matter Division of the EPS. Eds. J. L. de Segovia, F. Flores and F. Garzia-Moliner. 228 pages.	342

Physica Scripta Reprint Series

- RS1** **Perspectives in Condensed Matter Physics**, Trieste Lectures in Theoretical Physics, 20th Anniversary Workshop 1984. Published in *Physica Scripta* vol 32, 1985. 33 pages, price SEK 50 or USD 5.

RS2 **Nonlinear Waves in Plasmas**, Riso National Laboratory, Roskilde, Denmark, August 13–16, 1985. 135 pages, price SEK 270 or USD 40. Editor: H. L. Pécseli.

RS4 **Contributed papers to the 8th International Conference on Vacuum Ultraviolet Radiation Physics**, Lund, Sweden, August 4–8, 1986. 276 pages, price SEK 530 or USD 88. Editors: P.-O. Nilsson and J. Nordgren.

RS5 **Colloquium on Atomic Spectra and Oscillator Strengths for Astrophysics and Fusion Research**, Toledo, Ohio, August 11–13, 1986. 89 pages, price SEK 180 or USD 30. Editors: L. J. Curtis and L. S. Anderson.

RS6 **18th Conference of EGAS, European Group for Atomic Spectroscopy**, Marburg, Germany, July 8–11, 1986. 80 pages, price SEK 150 or USD 25. Editors: M. Elbel and H. Hühnermann.

RS7 **Relativistic Many-Body Problems**, Trieste, Italy, June 30–July 4, 1986. 240 pages, price SEK 480 or USD 80. Editor: I. Lindgren.

RS8 **Dynamical Screening and Spectroscopy of Surfaces**, Trieste, Italy, June 24–27, 1986. 170 pages, price SEK 340 or USD 57. Editors: M. Sunjić and S. Lundqvist.

RS9 **Quantum Chaos and Physics of Structure and Complexity**, Trieste, Italy, June 17–20 and September 1–5, 1986. 200 pages, price SEK 400 or USD 67. Editors: G. Casati, M. C. Gutzwiller, K.-E. Eriksson, H. Haaken and T. Arecchi.

RS10 **Environmental Physics – Atmospheric Aerosols**, Trieste, Italy, July 22–25, 1986. 112 pages, price SEK 220 or USD 40. Editors: H. Morinaga and E. Selin.

RS12 **19th Conference of EGAS, European Group for Atomic Spectroscopy**, NIHE Dublin, Ireland, July 14–17, 1987. 76 pages, price SEK 140 or USD 30. Editors: T. Andersen, E. T. Kennedy and G. O'Sullivan.

RS13 **Contributed papers to the 7th General Conference of the Condensed Matter Division of the EPS**, Pisa, April 7–10, 1987. 512 pages, price SEK 800 or USD 130. Editors: F. Bassani, G. Grossi, G. Pastori Parravicini and M. P. Tosi.

RS14 **The NORDITA-DIKU Conference on Vision**, Copenhagen, August 1986. 76 pages, price SEK 140 or USD 30. Editors: J. Hertz, P. Lennie, N. Graham and P. Johansen.

RS15 **Frontiers in Atomic Structure Calculations and Spectroscopy**, Lund, Sweden, May 25–26, 1988. 96 pages, price SEK 200 or USD 40. Editor: I. Martinson.

RS16 **Nonlinear Phenomena in Plasmas**, Roskilde, Denmark, August 8–12, 1988. 216 pages, price SEK 300 or USD 50. Editor: H. L. Pécseli.

RS17 **Contributed papers to the Conference on Soft Magnetic Materials 8**, Badgastein, Austria, September 1–4, 1987. 264 pages, price SEK 350 or USD 55. Editor: R. Wäppeling.

RS18 **Physics of Atomic Collisions**, Finse, Norway, May 8–11, 1989. 112 pages, price SEK 250 or USD 50. Editors: J. M. Hansteen and I. Singstad.

RS19 **Dusty Plasmas**, Trieste, Italy, June 14–21, 1991. 88 pages, price SEK 200 or USD 40. Editors: P. K. Shukla, U. de Angelis and L. Stenflo.

RS20 **Workshop on Physics with Penning Traps**, Lertorp, Sweden, June 29–July 1, 1991. 108 pages, price SEK 250 or USD 40. Editors: G. Bollen and C. Carlberg.

RS21 **The 2nd International Conference on Particle Production near Threshold**, Uppsala, Sweden, August 26–29, 1992. 212 pages, price SEK 300 or USD 50. Editors: C. Ekström, B. Höistad, A. Johansson and S. Kullander.

RS22 **Trends in Atomic Structure Calculation and Spectroscopy**, Lund, Sweden, March 3–4, 1993. 144 pages, price SEK 220 or USD 30. Editors: J. Carlsson, P. Jönsson and I. Martinson.

Order form: To The Royal Swedish Academy of Sciences, Physica Scripta, Box 50005, S-104 05 STOCKHOLM, Sweden

Please send me issues of:

RS1 RS2 RS4 RS5 RS6 RS7 RS8 RS9 RS10 RS12 RS13 RS14 RS15 RS16 RS17 RS18 RS19
RS20 RS21 RS22

Payment enclosed Please invoice Please charge my VISA MASTERCARD American Express card

Card number:

Expiry date:

Signature.....

Invoice address

Shipping address (if not invoice address)

Photonics	
Ultrafast nonlinear optics in GaAs/AlGaAs quantum wells. <i>J. M. Hvam, D. Birkedal, V. G. Lyssenko, J. Erland and C. B. Sørensen</i>	181
Modulation- and transmission-ellipsometric characterization of semiconductor heterostructures. <i>K. B. Ozanyan, T. Worren and O. Hunderi</i>	187
All-optical bistability in luminescence of thin CdS films. <i>B. Ullrich and T. Kobayashi</i>	191
Process technology	
Selective-area MOVPE for InP-based optoelectronic components. <i>P. Tidemand-Petersson, O. Albrektsen and J. Salzman</i>	194
Interdiffusion and phase formation during thermal processing of Co/Ti/Si(100) structures. <i>J. Cardenas, S. Hatzikonstantinidou, S.-L. Zhang, B. G. Svensson and C. S. Petersson</i>	198
Laser ablation deposition as a preparation method for electronic materials. <i>S. Leppävuori</i>	202
Growth of silicon carbide on (100) silicon substrates by molecular beam epitaxy. <i>V. M. Airaksinen, J. Kaitila, H. Niemi, J. Lahtinen and J. Saarilahti</i>	205
Growth and characterization of compositionally graded, relaxed $\text{Si}_{1-x}\text{Ge}_x$. <i>A. N. Larsen, J. L. Hansen, R. S. Jensen, S. Y. Shiryaev, P. R. Østergaard, J. Hartung, G. Davies, F. Jensen and J. W. Petersen</i>	208
RBS channeling spectroscopy of Ge implanted epitaxial $\text{Si}_{1-x}\text{Ge}_x$ layers. <i>J. Saarilahti, Z. Xia, H. Ronkainen, P. Kuivalainen and I. Suni</i>	212
As capping of MBE-grown compound semiconductors; novel opportunities to interface science and device fabrication. <i>J. K. Grepstad, H. Husby, A. Borg, B.-O. Finland, R. W. Bernstein and R. Nyholm</i>	216
Process optimisation and characterisation of PBT structures. <i>S. Hatzikonstantinidou, H.-E. Nilsson, C. Fröjdh and C. S. Petersson</i>	226
A flip chip process based on electroplated solder bumps. <i>J. Salonen and J. Salmi</i>	230
Low temperature silicon epitaxy in a single-wafer RTP reactor with microwave heating. <i>S.-L. Zhang and R. Buchta</i>	234
The Au/Si(111) system studied by optical second-harmonic generation. <i>K. Pedersen and P. Morgan</i>	238
Fabrication of nanostructures using MBE and MOVPE. <i>J. Ahopelto, H. K. Lipsanen, M. Sopanen, T. Koljonen, T. Tuomi, V. M. Airaksinen, J. Sinkkonen and E. Sirén</i>	241
Pulsed laser ablation deposition of CuInSe_2 and $\text{CuIn}_{1-x}\text{Ga}_x\text{Se}_2$ thin films. <i>J. Levoska, S. Leppävuori, F. Wang, O. Kusmartseva, A. E. Hill, E. Ahmed, R. D. Tomlinson and R. D. Pilkington</i>	244
A study of dual conductance response to carbon monoxide of CdS and α -SnWO ₄ thin films. <i>J. L. Solis, V. Golovanov, V. Lantto and S. Leppävuori</i>	248
Recent advances in theory	
A cluster approach for modelling of surface characteristics of stannic oxide. <i>T. S. Rantala, V. Lantto and T. Rantala</i>	252
The ionisation process of α particles in mesoscopic structures: simulation by Monte Carlo method. <i>K. Tarnay, F. Masszi, T. Kocsis, A. Poppe and L. Kiss</i>	256
A theoretically accurate mobility model for semiconductor device drift-diffusion simulation. <i>E. Velmre, A. Udal, T. Kocsis and F. Masszi</i>	263
Different methods of noise reduction in Monte Carlo simulations of a Schottky diode. <i>H.-E. Nilsson, U. Sannemo and C. S. Petersson</i>	268
Solid state materials	
Cobalt disilicide (CoSi_2) Schottky contacts to 6H-SiC. <i>N. Lundberg and M. Östling</i>	273
The Pt/Si(111) interface and the properties of thin Pt layers on Si. <i>P. Morgen, B. Jørgensen and J. Gordon</i>	278
SiC – a semiconductor for high-power, high-temperature and high-frequency devices. <i>E. Janzén, O. Kordina, A. Henry, W. M. Chen, N. T. Son, B. Monemar, E. Sörman, P. Bergman, C. I. Harris, R. Yakimova, M. Tuominen, A. O. Konstantinov, C. Hallin and C. Hemmingsson</i>	283
Thermal oxidation of n- and p-type 6H-silicon carbide. <i>C.-M. Zetterling and Östling</i>	291
Strain relaxation in epitaxial $\text{Si}_{1-x}\text{Ge}_x$ layers during some silicidation processes. <i>O. Nur, M. R. Sardela Jr., H. H. Radamson, M. Willander, G. V. Hansson and S. Hatzikonstantinidou</i>	294
An investigation of the stability of copper germanide thin films in the presence of Si and SiO ₂ . <i>J. P. Doyle, B. G. Svensson, M. O. Aboelfotoh and J. Hudner</i>	297
Properties of iron silicide contacts to n- and p-type silicon. <i>U. Erlesand and M. Östling</i>	300
List of participants	305

Order form

To The Royal Swedish Academy of Sciences
Physica Scripta
Box 50005
S-104 05 STOCKHOLM
Sweden

Please send issues of "16th Nordic Semiconductor Meeting", Laugarvatn, Iceland. Physica Scripta, Vol. T54, 1994, 312 pages. Price: SEK 468.

Payment enclosed Please invoice Please charge my VISA MASTERCARD American Express card

Card number:

Expiry date:

Shipping address (if not invoice address)

Name:

Address:

Contents

Preface	5
Defects	
Configurational instabilities at isoelectronic centres in silicon. <i>G. Davies</i>	7
Lithium-gold-related defect complexes in n-type silicon. <i>E. Ö. Sveinbjörnsson, S. Kristjánsson and H. P. Gíslason</i>	12
Conversion electron Mössbauer spectroscopy study of iron silicide films grown by MBE. <i>M. Fanciulli, G. Weyer, H. von Känel and N. Onda</i>	16
Deep centre photoluminescence in nitrogen doped ZnSe. <i>I. S. Hauksson, S. Y. Wang, J. Simpson, K. A. Prior, B. C. Cavenett, W. Liu and B. J. Skromme</i>	20
Excitation of the 4f-electron of Pr ³⁺ in GaAs : Pr and Al _{1-x} As : Pr. <i>P. L. Thee, Y. K. Yeo, R. L. Hengehold and G. S. Pomrenke</i>	24
Passivation of shallow and deep levels by lithium in GaAs. <i>T. Egilsson, B. Yang and H. P. Gíslason</i>	28
Effect of the interatomic Si-Si-potential on vacancy production during ion implantation of Si. <i>K. Nordlund, J. Keinonen and A. Kuronen</i>	34
Design and fabrication	
Surface morphology study of titanium silicide formed on polycrystalline silicon. <i>T. E. Karlin, W. Kaplan and S. Zhang</i>	38
A fine pattern GTO thyristor fabricated using a self-aligned process. <i>M. Bakowski, U. Gustafsson, H. Elderstig, W. Kaplan, P. Norlin and U. Wennström</i>	42
Application of III-V-semiconductor based heterojunction bipolar transistors towards multi-Gbit/s 4 : 1 multiplexer. <i>G. Schuppener, B. Willén, M. Mokhtari and H. Tenhunen</i>	46
Multichamber processor for small semiconductor laboratories – the first results. <i>R. Punkkinen, H. Ihantola, K. Jokinen, T. Kuusela, H. Arvela, H.-P. Hedman and V. Saikku</i>	51
Design optimization of negative charge pumps for IGBT driver ICs. <i>M. H. Åberg</i>	54
Electrical devices	
Processing and characterisation of an etched groove permeable base transistor on 6H-SiC. <i>C. Fröjd, G. Thungström, S. Hatzikonstantinidou, H.-E. Nilsson and C. S. Petersson</i>	56
The influence of emitter properties on the heat generation in SiC and Si PIN diodes under forward conduction. <i>O. Tornblad, B. Breitholtz, M. Östling and U. Lindefelt</i>	60
High-voltage silicon carbide rectifiers – results of experiments and simulation. <i>L. P. Ramberg, S. Savage, U. Gustafsson and A. Schöner</i>	65
Metallizations of InP based on transition metals. <i>T. Clausen and O. Leistikö</i>	68
Impedance field and microwave power generation in InP diodes. <i>V. Gružinskis, E. Starikov and P. Shiktorov</i>	71
Contacts to monocrystalline n- and p-type silicon by wafer bonding using cobalt disilicide. <i>G. Thungström, C. Fröjd, P. Svedberg and C. S. Petersson</i>	77
Low dimensional systems	
Magneto luminescence of As-grown InAs/InP quantum well islands. <i>H. Sigg, P. Christianen, R. Houdré and A. Rudra</i>	81
Initial strain relaxation and optical quality in lattice mismatched InGaAs/GaAs single quantum wells. <i>S. M. Wang and T. G. Andersson</i>	84
Electron correlation effects in quantum dots. <i>A. Matulis</i>	88
Enhancement of the g-factor and spin-density wave state in a confined 2DEG in the quantum Hall regime. <i>V. Guðmundsson and G. Pálsson</i>	92
Dissipation in the quantum Hall effect by transverse circulation of electrons. <i>E. B. Hansen</i>	96
Spin resonances determination of the effective g-factor of electrons in low dimensional (GaIn)As/InP structures. <i>B. Kowalski, P. Omling, B. K. Meyer, D. M. Hofmann, V. Härtle, F. Scholz and P. Sobkowicz</i>	100
Mesoscopic systems	
Characteristic potentials for mesoscopic rings threaded by an Aharonov–Bohm flux. <i>M. Büttiker</i>	104
Raman scattering from a circular quantum dot. <i>A. G. Mal'shukov, A. Brataas and K. A. Chao</i>	111
Local field calculation for a spherical semiconductor quantum dot with parabolic confinement. <i>O. Keller and T. Garm</i>	115
Resonant tunneling: from model Hamiltonian to modern electronic devices. <i>K. A. Chao, M. Willander and Yu. M. Galperin</i>	119
Aharonov–Bohm effect in a quantum ring with strong electron–electron correlations. <i>I. V. Krive, R. I. Shekhter, S. M. Girvin and M. Jonson</i>	123
Modelling of devices	
Unified capacitance modelling of MOSFETs. <i>O. G. Johannessen, T. A. Fjeldly and T. Ytterdal</i>	128
GaAs/AlGaAs quantum well infrared photodetector arrays for thermal imaging applications. <i>L. Lundqvist, J. Y. Andersson and J. Borglind</i>	131
A compact model for the cutoff frequency in high speed bipolar transistors. <i>M. Andersson, P. Kuivalainen, Z. Xia, H. Pohjonen and H. Ronkainen</i>	136
Extrinsic versus intrinsic models for FETs. <i>T. Ytterdal, T. A. Fjeldly and K. Lee</i>	139
The effect of using different transport models in computer simulations of the permeable-base transistor. <i>H.-E. Nilsson, U. Sannemo, A. Koel, F. Masszi and C. S. Petersson</i>	141
Monte Carlo simulation of hot carrier noise in short n ⁺ nn ⁺ diodes. <i>V. Gružinskis, E. Starikov and P. Shiktorov</i>	146
Effect of partial ionization and the characteristics of lateral power diamond MESFETs. <i>K. J. Grahn, P. Kuivalainen and S. Eränen</i>	151
Quantum corrections to the threshold voltage of short channel MOSFETs. <i>P. Kuivalainen</i>	154
Physical modelling of vertical DMOS power transistors for circuit simulation. <i>M. Andersson and P. Kuivalainen</i>	157
A semi-analytic model of the permeable base transistor. <i>H.-E. Nilsson, U. Sannemo and C. S. Petersson</i>	159
Optical devices	
Interferometric, low thermal mass IR-absorber for thermal infrared detectors. <i>P. Eriksson, J. Y. Andersson and G. Stemme</i>	165
UV-sensitive photodetectors based on metal-semiconductor contacts on 6H-SiC. <i>C. Fröjd, G. Thungström, H.-E. Nilsson and C. S. Petersson</i>	169
Large area GaInAsP and GaInP solar cells for space applications. <i>K. Smekalin, K. Tappura and J. Lammasniemi</i>	172
Fitting of the solar cell /V-curve to the two diode model. <i>A. Hovinen</i>	175
Development of materials for blue/green light emitters. <i>K. Rakennus, P. Uusimaa, K. Smekalin, P. Savolainen and M. Pessa</i>	177